UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THE RELATIONSHIP BETWEEN CERTIFICATION STATUS AND TEACHER EVALUATION: AN ANALYSIS IN A LARGE URBAN DISTRICT

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF EDUCATION

By

CRISTINA MOERSHEL
Norman, Oklahoma
2022

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/336882

THE RELATIONSHIP BETWEEN CERTIFICATION STATUS AND TEACHER EVALUATION: AN ANALYSIS IN A LARGE URBAN DISTRICT

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

BY THE COMMITTEE CONSISTING OF

Dr. Mirelsie Velazquez, Chair

Dr. Kate Raymond

Dr. Daniel Hamlin

Dr. Timothy Ford

© Copyright by CRISTINA MOERSHEL 2022
All Rights Reserved.

I would like to first dedicate this dissertation to all of my teachers and friends in education. It is our frustration with a broken system that led me to this dissertation. I hope that my continued work in educational research helps to find ways to support those who are in the classroom now. Thank you to my husband for supporting my coursework by taking care me and of our amazing children. When I started this journey, we had one brilliant baby and now here at the end, we are a family of five! Without your continued support, I may not have completed this journey. Thank you to my mother, aunt, cousins, and family members who continue to work in the public schools. We need great people with great hearts to continue working for our children. Thank you to my committee members who guided me through this process by letting me join summer writing groups and providing valuable feedback when I needed it most. Thank you for pushing me to keep going!

Table of Contents

List of Tables viii

Abstract x

Chapter One: Introduction 1

Evolving Expectations of Teachers 3

Teacher Shortage in Oklahoma 8

Effectiveness of Alternative Certification 10

The Pathways to State Certification 11

Developing an Evaluation Tool for Teacher Quality 14

The Purpose of Teacher Evaluation 15

Evaluation Models 16

Study Purpose 18

Limitations of the Study 20

Significance of the Study 21

Definitions of Terms 21

Assumptions 22

Summary 23

Chapter Two: Literature Review 25

Introduction 25

What is Teacher Quality? 27

Why Measure Input Quality? 28
On the Marzano Evaluation 30
Who Works in an Urban School District? 31
Why Do We Have a Teacher Shortage? 33
Teacher Preparation for Urban Schools 38
Deficit Thinking May Hinder Relationships 39
White Teachers' Feelings While Working in Urban Schools 44
Teacher-Student Relationships 46
How Can Teachers Build Relationships? 48
Definition & Measurement of Teacher Quality 50
Quality and Effectiveness 58
Effectiveness Through Standardized Assessments 60
Performance of Emergency-certified Teachers 62
What Makes a Good Evaluation? 63
The Evaluator 65
Marzano Focused Teacher Evaluation Model 67
Conclusion 69
Chapter Three: Methods 74
Introduction 74
Research Design 74
Data 77

The Evaluator 83

Aspects of the model 86

Independent and Control Variables 90

Data Analysis 90

Chapter 4: Results 94

The Data 94

Results 96

Chapter 5: Discussion 127

Summary of Key Findings 128

List of Tables

Table 1. Variable characteristics included in the final sample. 80

Table 2. The bivariate correlation of dependent variables and teachers' certification type 95

Table 3. OLS Regression examining the relationship between teachers' certification status and total evaluation score 96

Table 4. OLS regression models investigating the relationship between teacher certification and the critical content sub-score indicator 98

Table 5. OLS regression models investigating the relationship between teacher certification and the sub-score indicator, establishing, and acknowledging adherence to rules and procedures 100

Table 6. OLS regression investigating the association of teachers' total evaluation score and sub-score indicator, policies and procedures 102

Table 7. Robust ordinal regression investigating the association of teachers' certification and total evaluation score 104

Table 8. Robust ordinal regression results exploring the association of teachers' certification and sub-score indicator, critical content 106

Table 9. Robust ordinal results to investigate the association of teachers' certificate and sub-score indicator, rules and procedures 108

Table 10. Robust ordinal regression results that explore the association of teachers' certification and the sub-score indicator, policies and procedures 110

Table 11. OLS regression investigating the association between teachers' certification type and total evaluation score teachers who had scores in all three indicators 112

Table 12. OLS regression investigating the association of teachers' certification status and subcategory indicator, critical content, of teachers who had scores in all three indicators 114

Table 13. OLS regression results investigating the association of teachers' certification and subcategory indicator, rules and procedures when the teacher had scores in all three indicators 116

Table 14. OLS regression results investigating the association of teachers' certification and subcategory indicator, policies and procedures for the teacher had scores in all three indicators 118

Table 15. Results of OLS regression models investigating the relationship between teacher certification and total evaluation score for only new teachers. ... 120

Table 16. OLS regression table investigating the relationship between teacher certification and sub-score indicator, critical content, for new teachers only 122

Table 17. OLS regression table investigating the relationship between certification status and sub-score indicator, rules and procedures, for new teachers .. 124

Table 18. OLS regression investigating the relationship of new teacher certification and sub-score indicator, policies and procedures 126

Abstract

Teacher shortages are a growing problem across the United States. As one means of addressing this problem, policy makers have sought to loosen barriers to the profession by creating provisions for emergency certification. Emergency-certified teachers require no teaching experience, although they are required to have some experience working in the subject area. Research on emergency-certified teachers reports that these teachers have lower value-added measures, but this literature lacks information on how these teachers perform on teacher evaluations. In this study, the relationship between teacher evaluation scores and a teachers' certification status is investigated. Data are derived from a sample of teachers (n = 1285) in a large urban district. Results indicate small but significant associations between teacher certification and evaluation scores, with a difference in scores between -0.22 and -0.80. Furthermore, the results are consistent across different statistical models.

Chapter One: Introduction

As a high school math teacher, first in Texas and then in Oklahoma, for over 10 years, in two states, I became one of the teachers who moved up into leadership quickly and was one of the rare teachers with an undergraduate degree in education. Like teachers all across the United States, many of my colleagues entered the school sites with emergency certificates after receiving degrees in other fields. Because urban schools are hard to staff, the administrators had few prospects for critical positions, and many teachers left their school sites within three years. Many had no intention of staying for an extended time to hone their craft of teaching.

I have always taught in urban school settings where 100% of the students received free or reduced-price lunch. Students struggled to meet state standards before entering our classroom, but through support, encouragement, and scaffolded teaching, my colleagues and I worked together to help them pass state-mandated math assessments and demonstrate increased competency on standardized assessments. My students and I developed friendly, caring bonds by working together throughout the school year, but they knew that such bonds might not last. At the end of May, they could list the teachers whom they knew would be leaving. They noticed teachers packing up classrooms prematurely or being chronically absent during the last month of school. Near the end of the school year, the students would ask me if I was going to come back the next year, and they repeated this question almost daily as if they thought I had been lying when I told them I would return. They learned that teachers would leave them,

often without a goodbye. Their distrust in teachers affected the relationships that were or were not able to form throughout the school year. In any personal relationship, when people leave you, it is hard to trust again, and these students learned not to trust teachers after years of new teachers, or sometimes no classroom teacher at all due to a critical and growing teacher shortage. As a veteran teacher, it pained me to see students having to start over each year with new faculty who had to learn all about them, their families, and their hard work and accomplishments in previous school years I myself even began to wonder how long newly hired teachers would last, sometimes even doubting whether they would make it to the Christmas break.

As a teacher leader, I spent hours training new teachers on how to develop lesson plans, engage students, operate student information systems, and read reports. These new teachers most often joined our schools through alternative pathways, including Teach for America and emergency certification. Emergency-certified teachers were generally eager to learn how to teach the content; they loved the topics but lacked knowledge about how to scaffold instruction for learning or how to manage common classroom procedures. I loved teaching them the best practices I had learned and developed, but I knew that I would have to keep repeating the same process each school year due to the revolving door of teachers coming into the school. They were untrained in classroom management, and they did not know how to find the state standards or assess student learning. Especially during the COVID pandemic of 2020-2022, it was immensely fatiguing to train brand-new teachers who came in through alternative pathways.

The stress went home with me, and my family took on a lot of the pain I felt navigating the shift to online teaching while supporting new teachers who had little to no experience in the classroom.

Although I appreciated these new teachers, I also felt frustrated with their rapid departure and its effect on my school, my department, and my students. These conflicting feelings led me to this dissertation. I desire to find teachers of quality who will remain in urban schools for an extended time to develop strong teacher teams and supportive student and community relationships. I wondered why the state of Oklahoma would continuously place teachers in classrooms with little to no perceived support, knowing that it was statistically quite likely that they would need to be replaced in less than three years. I thought about the differences in teacher certification levels and pathways and how that might affect the quality of classroom teaching. Through my engagement with this struggle, I hope to provide insight for our policy makers about how the current certification practice in Oklahoma affects teaching quality and influences the level of frustration, disappointment, and loss felt by students and teachers in schools with higher rates of teacher turnover. My work addresses this question: Should Oklahoma policy makers adjust the qualifications for alternative teacher pathways to improve the quality of teaching in our students' classrooms, or are the alternatively certified teachers already doing an adequate job?

Evolving Expectations of Teachers

Historically, teachers have served various roles such as caregiver (Gasser, Grutter, Buholzer, & Wettstein, 2018), knowledge instructor (Grasha, 1994), and disciplinarian (Lewis & Lovegrove, 1987). The responsibilities have changed over time. Earlier in this country's history, teachers educated students of all ages in one-room schoolhouses, and students came from very similar cultural, religious, and economic backgrounds. Today, teachers tend to work with a narrow grade band and are often specialists in a single subject area. Scholars have routinely found that the work of teachers has become more complex (Amin, 2016; Friesen & Jardine, 2009). Their students vary widely in culture, family structure, socioeconomic status, and learning needs. Many teachers use online platforms to provide virtual instruction to students at home. In 1778, if teacher did not conform to their community's norms, parents and town leadership questioned the quality of the teacher and sent them back to the hiring agency (Fraser, 2001). During World War II, changing roles and expectations of our education system led to requirements for teachers to teach civics and American values. If teachers deviated from the school's plan, they were often criticized by parents. Teachers were once held accountable for advancing public norms ranging from knowing God to understanding virtue and safeguarding order and liberty (Scribner & Layton, 2003). Today, teachers are accountable for students' learning outcomes by way of subject matter standards and standardized assessment (Hallinger, Heck, & Murphy, 2014). Since the publication of *A Nation at Risk: The Imperative for Education Reform* in 1983, policy makers have increasingly adopted teacher accountability measures in an attempt to make teachers more responsible for

student outcomes. The culmination of this shift towards accountability came with the passage of the the No Child Left Behind Act of 2001 (P.L. 107-110, 20 U.S.C. § 6319, 2002), which required teachers to meet certain accountability standards, including adequate yearly progress, and to be evaluated based on student performance (Linn, Baker, & Betebenner, 2002). Teaching quality has become an integral part of accountability in education across the whole world (Hallinger et al., 2014). Teachers have been singled out as being responsible for the failure of students, even though most teachers are concerned about their students and want to provide the best education possible (Fraser, 2018; Fraser, 2001), and even though students' successes and failures are influenced by many factors beyond teachers' control.

Policy changes direct many of the expectations placed on teachers. Studies have been completed to determine the impact of No Child Left Behind Act of 2001 on student achievement. Many of these publications have blamed teachers for students' poor performance and target teacher quality as the major problem in schools. Critics of education blame teachers for many societal issues, including the erosion of American economic competitiveness and productivity, the decrease in academic achievement across the country, and daily conflicts in society (Ingersoll, 2004). Research suggests that teacher quality has a strong effect on students' outcomes (Chett, Friedman, & Rockoff, 2014a, 2014b; Hanushek, 2011; Rockoff, 2004). This concern with teacher quality has created new certification and accountability measures that teachers must now concern themselves with to become certified classroom teachers. In such an environment, it is important that

policy makers understand what makes a teacher effective and how to measure effectiveness.

Many potential teachers seek out educational development programs through universities to cultivate their craft using a traditional teacher pathway. The traditional teacher path includes coursework in learning, assessment, classroom management, exceptional learners, literacy, and inquiry-based practices (Feuer, Floden, Chudowsky, & Ahn, 2013). These courses provide a foundation for teachers as they navigate the 21st-century classroom. Those who enter through the traditional teacher path participate in several semesters of peer observations and field experience to develop classroom management strategies and practice teaching with certified instructors who provide support and feedback (Darling-Hammond, 2010). These teachers enter the classroom with specialized education that allows them to begin their careers with suitable preparation. However, not all new teachers are graduates of teacher preparation programs; novice professionals often fill the vacancies through alternative routes that include no formal training (Laczko-Kerr & Berliner, 2002). The number of bachelor's, master's, and doctoral degrees in education awarded U.S. colleges and universities fell from 289,282 in 2009 to 244,851 in 2016, a decrease of 15.4% (Garcia & Weiss, 2019b). Likewise, the number of people completing teacher preparation programs fell during the same time period from 217,400 to 157,901, a decrease of 27.4%, reducing the pool of potential teachers (Garcia & Weiss, 2019b). On the other hand, during this time, the number of people enrolled in an alternative teacher preparation not related to an institution of higher education increased by 71%

from 38,595 to 66,173. Clearly, more potential teachers are enrolling in alternative preparation programs. But completion rates also have changed in similar ways. For alternative certifications, the number of participants completing their program decreased by 22.9 percent from 2009 to 2016, but the completion rate among participants in traditional pathways decreased more sharply, by 44.6% (Garcia & Weiss, 2019b). In response to the declining enrollment in traditional teacher certification programs and declining completion rates, many states, including Oklahoma, have modified certification requirements to reduce perceived barriers to entry. The pathways that potential teachers use are changing. Multiple certification paths exist, and their requirements vary from state to state (National Research Council, 2010).

For example, alternative pathways include Teach for America (TFA). TFA is a non-traditional pathway to teaching. TFA recruit students from elite colleges to serve as classroom teachers for two years in low-income urban schools that are hard to staff (Teach for America, 2022). TFA teachers participate in an intensive summer training program before entering the classroom, and TFA provides mentors and continued education to support the certification process (Teach for America, 2022). Troops to Teachers is a federally funded program that supports former military service members in obtaining certification and employment (OSDE, 2022). Established in 1993, the program provides financial incentives, including a $10,000 bonus for working in a high-needs school (OSDE, 2022). Pathways also exist for Career Development paraprofessionals, CareerTech instructors, and Four-Year-Olds and Younger certification (OSDE, 2022).

Emergency-certified teachers are able to enter the classroom with no prior teaching experience, and they do not have to be signed up for an alternative teaching program at the time the license is issued (OSDE, 2022).

Teacher Shortage in Oklahoma

Oklahoma is emblematic of many of the trends observed in teaching profession in recent years. Teachers in Oklahoma have faced a variety of critical issues that have pushed them out of the classroom at an alarming rate. Teachers have cited lack of respect and low pay as major contributing factors to their departure (Cullen, 2018). According to a 2017 survey by the Oklahoma State School Boards Association (OSSBA), superintendents reported starting the year with 536 teaching vacancies and 1430 emergency certificates. The shortage of qualified teachers caused by high attrition has affected the lives of thousands of students by reducing the number of courses offered at each school site and increasing class sizes (OSSBA, 2017). School districts have been unable to fill vacancies and have taken measures to reduce the need to fill these positions. Due to the high turnover, teachers hired to fill these vacancies are often new to the profession and may lack full certification and the expertise to develop critical skills in the classroom, such as classroom management, the ability to differentiate (i.e., create custom learning, for individual student), and competency in a subject area.

In September 2015, the Oklahoma State Department of Education (OSDE) created a Teacher Shortage Task Force to create actionable solutions to address the crisis. In its preliminary report, the team focused on reducing the state's

barriers to teacher certification. For example, it was often difficult for teachers from other states to transfer their teaching certificates, because they were required to have five years of experience in another state (Teacher Shortage Task Force, 2015). Additional recommendations by the task force included subsidizing fees for tests, reducing the requirements for entry into the alternative teacher program by allowing work experience to replace the required 2.5 GPA requirement, and a recruitment incentive program (Teacher Shortage Task Force, 2015). In response, OSDE in 2016 amended its requirements for alternative certificates in the hope of filling the increasing vacancies. Under the amended requirements, alternative certificates now allow teachers to enter the classroom who might have experience in a specific field other than teaching, which changes the traditional demographic of teachers from specialists in pedagogy to those with related field experience (Kirby, Darling-Hammond, & Hudson, 1989). OSDE's purpose in amending the requirements was to create an attainable path towards certification by allowing interested parties to follow an alternative certification program (Teacher Shortage Task Force, 2015). By allowing other professionals to enter the teaching field directly, without participating in college coursework in education, these individuals would be able to receive a paycheck and simultaneously attain some pedagogical training. Stakeholders throughout Oklahoma felt that if someone had to stop working to take two years of coursework, it might deter them from making a career change to teaching (Teacher Shortage Task Force, 2015). This benefits teachers because they can apply their job experiences to the classroom immediately instead of pausing to go to school (Kamenetz, 2014). An objective of

the task force was to eliminate barriers, and it did so by streamlining training, employment, and steps to certification. Proponents of alternative certification believed that teachers entering the classroom with applicable field experience would be just as successful as teachers who participated in years of pedagogical training. In other words, knowing the content would make them good teachers.

Effectiveness of Alternative Certification

How will Oklahoma know if the policy changes made to the alternative teacher pathways actually ameliorate the teacher shortage and actually place high-quality teachers in front of our students? The state of Oklahoma has not defined ways to measure the success of the policy changes to the alternative teacher program, and the requirements vary depending on the program model (Bowling & Ball, 2018).

Effective alternative certification programs should include academic selectivity, strong content knowledge, a course sequence that prioritizes coursework that will directly aid in developing a good teacher, and support that includes being paired with a mentor (Walsh & Jacob, 2007). In a study of 49 different alternative certification programs, Walsh & Jacob (2007) found that none of the programs met all four characteristics. This included programs operated by universities, districts, and private organizations.

Potentially, the state may use the continued retention of teachers and the assessment of comparable levels of teacher quality between alternatively certified and traditional certified teachers to measure the success of alternative routes

towards certification. Focus groups composed of district leaders throughout the state may also provide an appropriate analysis of the alternative teacher programs' progress in executing alternative certification and help determine which aspects work and which do not (Gardner & Brindis, 2017).

Oklahoma's current alternative certification process requires potential educators to complete two courses on classroom management and subject-specific pedagogical principles within a three-year period after employment begins (OSDE, 2018). Teachers who enter through alternative pathways are now required to have a mentor through a teacher induction program. Each school site holds responsibility for assigning mentors, who are asked to volunteer time each month to work with new educators. Teachers who are working with an alternative certificate are left on their own to enroll in the additional courses through an accredited university and to pay for required testing after they are already in the classroom. Although the number of alternative teaching certificates has increased since the policy change, an insufficient amount of research has been conducted to assess the quality of daily teaching being performed in classrooms by these teachers.

The Pathways to State Certification

States' teaching certification programs exist to ensure that teachers meet the minimum standards of competency in a subject area and general teaching practices. In Oklahoma, various pathways towards teacher certification exist. Traditionally, teachers obtain a bachelor's degree from an accredited four-year teacher preparation program that includes classroom observation hours (NCES,

2018). Regulations require that these teachers pass certification exams before obtaining a license through the state of Oklahoma (NCES, 2018). The alternative teacher certification pathway allows those who hold a bachelor's degree in any subject area to teach in a classroom while working toward completing a teacher preparation program that includes state certification exams (NCES, 2018). Teachers who pursue an alternative certification route are required to complete 6 to 18 college credit hours in professional education or 90 to 279 clock hours of professional development. Teach for America (TFA) recruits college graduates from prestigious institutions by offering extended support and financial incentives for those who complete the program. TFA recruits spend two years teaching in low-income schools with minoritized students. They are expected to complete a brief but intensive summer training session before entering the classroom (Teach for America, 2020). If a school district has exhausted all efforts to employ a certified teacher, emergency certification allows anyone with a bachelor's degree to become a classroom teacher without having completed a traditional teacher preparation program or being enrolled in an alternative teacher preparation program (Oklahoma State Department of Education, 2021). An emergency certificate lasts one year, and teachers must reapply each year. In Oklahoma, a person may be issued three emergency certificates if they have not successfully completed required competency examinations but have been working toward certification (Oklahoma State Department of Education, 2021).

Despite reducing barriers to certification, Oklahoma continues to face a critical teacher shortage due to low pay and the negative stigma associated with

the profession (Burkhardt, 2017). OSSBA's 2017 survey outlined the severity of Oklahoma's teacher shortage, revealing 536 vacancies, 480 eliminated positions, and between 1430 and 2000 emergency certificates in place or filed for processing (OSSBA, 2017). In 2019, Oklahoma schools saw a similar trend, with 596 vacancies and a reduction of 207 teaching positions when the school year started (OSSBA, 2019). During the first half of the 2019-2020 school year, 2801 emergency teaching certificates were issued, and in 2021-22, the figure was 2673 (Palmer, 2021). In June 2017, state superintendents also reported more difficulty filling positions than in previous years, with the most difficult areas being math, science, and special education (OSSBA, 2017).

The Teacher Shortage Task Force, first convened in 2015, has continued to work on recruiting, retaining, and rewarding educators (OSDE, 2018). In 2017, the task force reported that 38,105 of the state's approximately 695,000 students were being taught by emergency-certified teachers. In May 2017, Burkhardt found that 50,000 students were being taught by alternatively certified teachers who had no previous teaching experience. A large proportion of teachers are emergency-certified throughout the United States, and the number continues to increase (Learning Policy Institute, 2018). The brunt of the U.S. teacher shortage is felt in urban schools with high minority enrollment (Darling-Hammond, 2010; Rahman, Fox, Ikoma, & Gray, 2017; U.S. Department of Education, 2016).

Not having enough fully certified teachers is only one of several factors contributing to Oklahoma's education crisis. Oklahoma's certified teachers are leaving the state at alarming rates to pursue teaching careers in other states

(Cullen, 2018). Only 81.8 % of teachers stay beyond one year, 68.7% beyond two years, and 53.0% beyond five years (Teacher Shortage Task Force, 2018). The Oklahoma Education Coalition (OEC), which works to inform citizens about education legislation, finds that 11% of teachers leave Oklahoma each year and that, between 2010 and 2015, the state experienced a net loss of teachers (OEC, 2021). Preliminary data from Cullen (2018) shows that 80% of teachers who left Oklahoma for neighboring states did so for financial reasons. The teacher shortage has led to an increasing number of vacancies, outpacing the number of teachers completing traditional preparation programs (Bowling & Ball, 2018; National Research Council, 2010). Oklahoma currently does not have enough traditionally certified teachers to fill every classroom in every school district. The disproportionate number of vacancies has created an increasing demand for teachers who are willing to take an alternative pathway into the classroom.

Developing an Evaluation Tool for Teacher Quality

President Obama's Race to the Top program, launched in 2009, used grants to persuade states to modify or develop teacher evaluation tools and thereby improve student learning. Although Oklahoma's application for Race to the Top funding was unsuccessful, the modifications of teacher evaluation proposed in the state's application became become part of Oklahoma law in 2010 as a way to improve student outcomes (Oklahoma State Courts Network, 2012; Oklahoma State Department of Education, 2014). In 2011, the Oklahoma Teacher Leader Effectiveness Law (TLE) mandated each school district to choose a

teacher evaluation framework (Oklahoma State Court Network, 2012). Asserting that teacher effectiveness is one of the most important factors in student success, the OSDE in 2019 implemented an evaluation program that provided professional development to address weaknesses and provide teachers with ongoing feedback (Oklahoma State Department of Education, 2019). However, researchers have raised concerns about the effectiveness of teacher evaluations because many of the tools focus on how schools and districts are organized rather than on teachers' classroom practices and students' learning (Hallinger & Heck, 2014).

The Purpose of Teacher Evaluation

Teacher evaluation has been presented in various forms over the years. Some evaluation programs include value-added models (VAM) alongside lesson observations that might include student learning targets. VAMs estimate a teacher's contribution to a student's growth over several years. Generally, these models account for prior student growth. However, the purpose of teacher evaluation has remained the same: to assess a teacher's practice in a formative way to improve student learning (Ford, Van Sickle, Clark, Fazio-Brunson, & Schween, 2017). Teachers often improve practice when their motivation is intrinsic; extrinsic motivators, such as incentive pay, cause very little increase in a teacher's motivation to improve student learning or to remain in the classroom (Lavigne & Good, 2014; Marsh, Springer, McCaffrey, Yuan, & Epstein, 2011). Intrinsic incentives include building relationships, increasing skills through professional development, and spending time collaborating with others and

reflecting on processes. Evaluations that provide teachers with specific, measurable, and formative analysis attempt to provide teachers with these critical skills. When teacher evaluations do not meet the needs for intrinsic motivation, they create a stressful environment that demoralizes teachers (Ford et al., 2017). Teachers need interactions with other teachers to fully understand the expectations of teacher evaluations. If school districts fail to provide support that includes collaboration, discussion, and reflection, the teacher evaluation cannot succeed. When crafting evaluation tools, therefore, it important to understand whether the proposed evaluation program has measures and tools that motivate teachers to improve and that drive their professional development.

Evaluation Models

In Oklahoma, school districts must implement an approved evaluation model to meet the TLE requirements. The two most utilized models include Marzano and the Tulsa Model. The Tulsa Model was created in 2010 by Tulsa Public Schools (TPS) to increase student achievement by improving instructional practices through feedback and analysis. The district gathered national experts, teachers, and administrators to create and develop rubrics that would correlate with professional competencies and best practices in the classroom (TPS, 2015). After its first year of implementation, the results were evaluated by the University of Wisconsin, and indicators included in the model were directly correlated with increasing student achievement (TPS, 2015). The Tulsa Model was developed for use in Oklahoma and has been used in 499 of the state's 740 school districts (TPS, 2022). The Marzano Focused Teacher Evaluation Model was developed by

Robert Marzano over the last two decades, based on 5000 studies that have identified specific elements correlated with students' academic achievement (Marzano, Toth, & Schooling, 2012). The Marzano Center has validated results of its model by examining student achievement and using various state assessments and teachers' overall proficiency scores (Marzano et al., 2012). However, the evaluation model does not include value-added measures as a measure of teacher evaluation itself. Instead, the model focuses on specific elements that a teacher has control over and the ability to improve. The elements include planning standards-based units, helping students process new content, and promoting teacher leadership and collaboration, which are key factors in teacher motivation (Ford et al., 2017). The Marzano Focused Teacher Evaluation Model is formative in nature, providing teachers with the space to use their results to create specific and personal goals, thereby enabling intrinsic motivations to improve. The model fulfills teachers' desire, revealed in previous research, for dialogue between teachers and evaluators, peer observations, and evaluation debriefing processes.

Prior Literature and Theory

Differences in quality between traditionally certified teachers and alternatively certified teachers may exist due to the relaxed requirements for alternative teaching certification. Current literature finds that differences do exist between certification types when measuring teacher quality through value-added methods (Walsh & O'Tracy, 2005). However, the magnitude of the difference has been questioned in several studies (Walsh & O'Tracy, 2005). Darling-Hammond, Berry, and Thoreson (2001) questioned a previous study's finding that students

fared the same at the secondary level no matter what kind of certification was held by the teacher. The researchers then countered these findings by identifying a significant effect of teacher certification level on students' success in mathematics, but not in all subjects or grade levels. A large study in New York also found that some differences existed between certification types and teacher quality (Kane, Rockoff & Staiger, 2008). Denton and Lacina (1984) found that a teacher's education coursework significantly affected student achievement; this evidence is cited by Darling-Hammond et al. (2001), who also found that teacher education was correlated with student learning. Much of the current literature is grounded in the 2001 work of Darling-Hammonds' et al., building on the idea that the teacher shortage is caused by more than a growth in the population. However, the literature fails to use measures of teacher quality that can be controlled directly by a teacher's actions, and the literature has little evidence linking a teachers' certification level and the teacher evaluation, a measure of quality. Following this line of research, my work explores the connections between certification type and teacher quality within Oklahoma's challenging education environment.

Study Purpose

This study is based on the hypothesis that alternatively certified teachers have characteristics that affect the quality of teaching they provide in urban schools. We know that urban school districts have been increasing the number of teachers who hold alternative certificates without formal education training (Cullen, 2018; Darling-Hammond et al., 2001). By examining evaluations of all

teachers in one district, this study investigates the potential differences in teacher quality in urban schools between alternative and traditionally certified teachers as based on formal evaluation scores assigned by principals. The study examines the characteristics of successful teachers as identified by a consistent evaluation method and compares scores to identify differences in teacher quality as appraised by school principals.

This study investigates the potential differences in teacher quality among teachers certified through traditional, alternative, and emergency certifications in an urban school district. Traditionally certified teachers have participated in a state-approved teacher education program. Alternative certification paths include those who have not completed a state-approved education program but hold at least a bachelor's degree. Non-traditional paths include American Board for Certification of Teacher Excellence and Teach for America. Emergency certificates include those who might be working towards a bachelor's degree and will be participating in an alternative certification process. The quantitative research will explore and examine aspects of teacher quality and the predictive factors that may influence differences in practice based on certification by engaging in a substantive literature review. The directional hypothesis to be tested is that teachers who hold an alternative or emergency certificate have lower principal-conducted evaluation score on the Marzano Focused Teacher Evaluation Model (MFTEM) than fully certified teachers who were educated through traditional routes. Results of the study will potentially indicate whether

certification type is associated with teaching quality as judged by principal evaluators.

H^1: A difference exists in teacher quality between teachers who hold traditional teacher certificates and those who hold alternative and emergency certificates.

H^2: A difference in teacher quality exists between first-year teachers who hold a traditional teacher certificate and those who hold an alternative and emergency teaching certificate.

Through an analysis of existing evaluation data, the study assesses the quality of teaching provided by alternatively certified teachers compared to traditionally certified teachers as determined by their evaluating administrator. An investigation of evaluation data indicators will shed light on the quality of teaching provided by alternatively certified teachers and determine whether there is a need to modify Oklahoma's current licensing procedures.

Limitations of the Study

As with all research projects, this study's design and sampling process have limitations. Due to the ongoing pandemic, teacher evaluations have been limited and put on hold throughout Oklahoma. The most recent teacher evaluation data are for the school year 2018-2019. Due to the variety of evaluation tools offered for use in the state of Oklahoma, only school districts that use the Marzano evaluation tool were part of the initial data-gathering. From there, the school district selected was the largest in the state that utilizes the Marzano tool. Smaller districts were not selected due to the small sample size. In future studies,

it would be important to collect data from multiple suburban and rural districts to determine whether the findings are specific to urban schools or generalizable to all of Oklahoma.

Significance of the Study

This study identifies the differences in teacher evaluation scores, a measure of quality, and a teachers' certification in an urban school district. The results of this study help fill gaps in the existing literature. The results may also help principals and other administrators understand the differences between certification levels and inform their decision-making regarding funding allocations and professional development opportunities. Additionally, principals are ostensibly instructional leaders who have the qualifications and ability to make informed and expert judgments about the quality of teaching and learning that occurs within their school sites. This study's findings may provide principals with better information about the level and kind of discernment they give to their observation and rating of teachers' work. Moreover, by shedding light on potential differences between teachers with different certifications, this study also could provide valuable information to federal and state officials reviewing legislative priorities. The findings also may impact the development of new legislation on teacher certification and/or principal practice.

Definitions of Terms

Emergency-certified teacher. A person who has not completed and is not currently enrolled in a teacher preparation program and has not passed content-specific state teacher exams.

Alternatively certified teacher. This designation includes those who originally may have been issued an emergency-certificate and those seeking licensing without having completed a traditional university teacher preparation program that includes observation and student teaching practicum hours. These teachers are working toward a standard certification and have enrolled in a teacher preparation program.

Traditionally certified teacher. This status includes two groups: teachers who have completed an alternative certificate program and now hold a full certificate, and teachers who attended a traditional university-based program that includes observation and student teaching practicum hours and hold a bachelor's degree from an accredited college. In Oklahoma, a traditionally certified teacher must pass the OGET, OSAT, and either the OPTE or PPAT. They hold a standard certificate for five years, issued by the Oklahoma State Department of Education.

Quality teaching. As designated by the state of Oklahoma, quality teaching is indicated by a score of three or greater on a five-point teacher evaluation tool.

Assumptions

This study assumes that the Marzano Focused Teacher Evaluation Model evaluation tool is accurate and that evaluator scores have been calibrated through district training. The school district being investigated in this research hosts quarterly walkthroughs with all evaluators. At these sessions, groups of evaluators

enter the same classroom at the same time to observe the same setting. The evaluators might be school site principals or assistant principals who have completed pre-requisite training sessions on the Marzano evaluation. They evaluate the teacher, leave the classroom, and then debrief to calibrate scores with a trainer who helps them describe what they saw and how that score relates to the instrument being used. These evaluators go through this process again in another classroom and repeat the process regularly to ensure equity and fairness in scoring practices across the district. This boosts the interrater reliability of the Marzano Focused Teacher Evaluation Model as used by different evaluators across the school district. It is also assumed that all teachers who receive evaluations have been equally notified about the evaluation and have had adequate training, modeling, and preparation for the evaluation process.

Summary

The teacher shortage has created an increasing number of vacancies that outpace the number of teachers entering the classroom (Bowling & Ball, 2018; National Research Council, 2010). In Oklahoma, the number of teachers completing traditional teacher preparation programs continues to decline while the number of teachers with alternative or emergency certificates experiences significant growth. Oklahoma is an important site for this research because the state has teacher shortages similar to those in surrounding states but is experiencing greater difficulty in retaining fully certified teachers as legislation decreases spending and increases regulations on school districts (Maiden & Reynolds, 2019). This quantitative study will utilize teacher evaluation data in

one urban district that is experiencing a teacher shortage and an increase in alternatively certified teachers.

Research tells us that teacher quality matters in all school settings but especially in urban school districts, where teachers are often the least qualified for their positions (Ingersoll, 2004; Jacob, 2007). This project will attempt to shed light on any differences in teacher quality based on a teacher's certification type to determine whether current Oklahoma teacher certification policy should be modified to improve educational outcomes. The review of the literature defines teacher quality, the technical specifications of the Marzano Focused Teacher Evaluation Model, the need for highly qualified teachers within urban school settings, and the role of school principals as instructional leaders.

Chapter Two: Literature Review

Current literature often uses value-added measures to identify and measure teacher quality. However, teacher quality is based on much more than a standardized assessment. Current literature fails to use current teacher evaluation models as a way to measure teacher quality, and it fails to address the association between a teachers' certification level.

Introduction

This chapter will review several bodies of literature that inform this study. The review of literature covers current definitions of teacher quality and how it is currently measured in schools throughout the United States. It also describes the Marzano Focused Teacher Evaluation Model (MFTEM) tool and explains its importance and how it has been adapted to meet the changing needs of education. In addition, the literature review evaluates current research that describes the type of teachers currently working in urban schools and their level of preparation in the areas of quality that particularly matter for the success of urban students.

For learning to occur, students must actively participate in learning and have a supportive environment, opportunities to learn, and good teaching (Fenstermacher & Richardson, 2005). Without all of these elements, a teacher could provide instruction of the highest quality, but the student would be unable to receive and process the information or acquire the skills of higher-order critical reasoning. Literature suggests that students who live in high-poverty areas often need the best teachers, yet for various reasons, they are given some of the least qualified teachers (Redding & Nguyen, 2020; Lankford, Loeb & Wyckoff,

2002). When parents complain about low-quality teachers, those teachers are likely to be transferred to schools with a large proportion of minoritized students (Bridges, 1996). Schools in such communities may serve a large population of English language learners (ELL) that requires additional services yet be unable to fill vacancies for ELL services (Guin, 2004). Urban communities and families often value education and hold teachers to standards that are unfamiliar to the teachers who serve them, which causes stress and requires additional training in cultural awareness and social justice. However, because of the shortage of traditionally educated teachers and the prevalence of alternatively certified teachers, these communities' teaching positions are often filled by teachers who have had no training in social justice or do not know how to work with students living with trauma and issues stemming from generational poverty. Although teachers accept these positions, they are often ill-equipped to provide a quality education to these students. It is important to consider what factors contribute to the success of teachers in urban environments when they are unfamiliar with the community norms and values.

The 2003 report *No Dream Denied*, published by the National Commission on Teaching and America's Future, found that half of new teachers will leave the profession by their fifth year (Jacob, 2007). If this trend continues, by 2025 Oklahoma will no longer have enough quality educators willing to teach in urban districts. But before delving into this problem, it is important to understand what defines a quality teacher, the measures that might be used to evaluate teaching quality, and how teacher quality impacts student learning so that

principals can inspire growth and support teachers through accurate, quality evaluation tools.

What is Teacher Quality?

Many studies that attempt to determine teacher quality do so by examining student performance on standardized testing. When teacher quality is measured in this way, little difference is found between traditionally and alternatively certified teachers. Hanushek's study found that ineffective or low-quality teachers harm student learning and, after several years of ineffective teaching, students fall behind their peers (Hanushek, 2009). Students of low-quality teachers only show one-half of a grade level in gains compared to students of quality teachers (Hanushek, 2009). However, if a student has an above-average teacher for four or five years, the student will close the average learning gap (Hanushek, 2009). Rice (2003) found that teacher certification, a possible measure of quality teaching, did not matter at the secondary level in courses (except mathematics) in which standardized assessments were used as an output measure. No differences were found in ELA teachers, but there were differences in math. Rice also found that, at the elementary level, teacher certification produced no significant differences in results of standardized assessments. Moreover, Rice's study showed that teaching experience mattered only in the first few years of teaching, which paralleled findings by Goe (2007), who noted that after the first four or five years, teachers no longer contributed to additional student achievement scores beyond what they previously had measured.

Wilson, Floden, and Ferini-Mundy (2001) identified gaps in the research concerning teacher preparation programs. They noted that, to study teacher preparation programs, researchers must examine effects over time and study relationships between specific components of specific programs instead of broad comparisons of preparation programs. Yet current research on teacher preparation as a measured input has failed to address these 2001 findings and continues to broadly discuss teacher training programs. The current literature suggests that having a quality teacher might make no difference in student outcomes, a measure of teacher quality. However, the research mostly covers only the subjects of English and mathematics, and the methods of calculating quality demonstrate significant room for error and interpretation.

Qu and Becker (2003) conducted a meta-analysis involving teacher certification type. Through an evaluation of previous research, they found that traditionally certified teachers generally performed at the same level as alternatively certified teachers, but outperformed emergency-certified teachers. They also noted that research involving teacher certification and teacher quality or performance often reported insufficient information for the findings to be used to compare studies (Qu & Becker, 2003). For instance, several reports failed to provide sufficient information about the alternative programs' requirements or how long teachers were required to attend the program before earning a full teaching license.

Why Measure Input Quality?

Studies have shown that teacher quality could be a contributing factor in student achievement, but researchers disagree on how to measure quality. These studies have failed to show significant differences between traditionally and alternatively certified teachers or any improvement in teacher quality after the fifth year of teaching when equating teacher quality with student achievement on standardized tests. Current research models focus only on math and english, as these are the subject areas that utilize standardized assessments throughout a student's schooling. Thus, these models predict effectiveness in only two content areas and fail to evaluate the effectiveness of teaching in all other courses offered by schools in Oklahoma.

Instead of focusing on one output, Marzano and Toth (2017) focus on several: the processes and procedures teachers work through every day to allow students to grow within every content area. Marzano's research on teacher quality has shifted over time due to changes in policy and deep reflection on the most important aspects of teaching. The research focuses on student learning within the classroom as a measure of teacher quality but addresses it from various practical, measurable perspectives (Marzano & Toth, 2017). These include how a teacher asks deep probing questions and redirects students to think about the task at hand instead of just providing answers. The research focuses on all classroom teachers and best practices, including specific, direct, and measurable goals in the area of professional development, instead of simply assigning a teacher a number based on one test at the end of the year. The Marzano team prides itself on conducting

research to inform multiple assessments that allow teachers to adjust their practices, learn from others, and grow professionally.

On the Marzano Evaluation

Using a sample of over 12,000 teachers in Florida, Basileo and Toth (2019) employ teacher value-added measures and the MFTEM to assess the model's predictability. Through a correlation of the instructional frameworks, they found that teacher observations were the strongest predictor of the value-added measure (Basileo & Toth, 2019). Although Strong et al. (2007) did not specifically utilize the same evaluation tool, the qualities observed closely align with the MFTEM and identify high-quality, effective leaders based on the same elements. Kane and Staiger (2012) also suggest that accurate and helpful evaluations require multiple measures to compensate for the imperfections of the individual indicators. These are qualities found in the MFTEM.

Currently, the MFTEM is being used in fewer than 40 school districts in Oklahoma. This may be due to the costs associated with the program. MFTEM uses an online platform known as iObservation to collect, manage, and report data from teacher evaluations, observations, and informal walkthroughs. The platform also offers additional support for teachers through discussion boards and video lessons that target specific classroom management and instructional practices that are valuable for individual growth. For a full day of training on iObservation, a school district would pay approximately $4500 for only 35-40 people. Virtual training costs $199 per participant, and professional development on the

framework costs $4,500, with annual fees costing much more than the Tulsa Model.

Hazi (2014) claims that no link between teacher evaluations and teacher quality exists using the MFTEM. She also asserts that Marzano's marketing preys upon vulnerable school districts. Under new regulations regarding teaching quality, some districts are forced to fire low-performing teachers. Hoping to protect themselves against unjust termination lawsuits, such districts may be lured into using Marzano's program due to its claims of validity, reliability, and objectivity (Hazi, 2014).

It is important to understand why some may view the Marzano evaluation tool as an invalid measure so that researchers might take precautionary measures to prevent error in future studies. However, minimal literature exists to suggest that the MFTEM is ineffective when used with fidelity. Marzano has been able to weave together many of the agreed-upon aspects of teacher effectiveness into one specific, direct evaluation model. The MFTEM focuses on teacher and student growth and provides the school district with ample training opportunities in person and online to ensure successful program implementation.

Who Works in an Urban School District?

Teachers are leaving urban school districts at an alarming rate compared to their suburban counterparts, and urban school districts risk having underqualified teachers fill the urgent vacancies (Freedman & Appleman, 2008; Ingersoll, 2001, 2003; Jacob, 2007; Loeb, Darling-Hammond, & Luczak, 2005; Sleeter, 2001). Forty-six percent of new teachers will leave the profession by the

end of five years (Ingersoll, 2002), and nationally, 24% of teachers leave after one year of teaching (Freedman & Appleman, 2008). The overall teacher turnover rate in 2001 was 13.2% in all schools and 14.4% in urban, high poverty schools (Ingersoll, 2001). This trend has continued, with departures by 40% of new teachers with degrees and 54% of those without full certification (Reeding & Nguyen, 2020). Cochran-Smith (2004) notes that fast-track entry will not resolve the teacher shortage problem. The problem is multidimensional, and the desire to teach children is not enough to keep teachers in the classroom.

Ingersoll (2012) describes the first-year teacher's experience as a "sink or swim" situation due to isolation and lack of preparation. In urban school districts, predominantly white teachers typically serve students of another cultural background. White teachers often come into the classroom with stereotypic beliefs about students' abilities (Leland & Harste, 2005; Sleeter, 2001; Schultz, Neyhart, & Reck, 1996), leaving them unprepared for the students and the urban settings (Lankford, Loeb & Wyckoff, 2002; Sleeter, 2001). New teachers also tend to be assigned to the most challenging classrooms and schools, which have the least-desired positions and therefore the most vacancies (Ingersoll, 2002; Steele, Pepper, Springer, & Lockwood, 2015). Steele et al. (2015) analyzed seven years of achievement data in an urban school district, determining that mobility rates were not determined by a teacher's value-added effectiveness and that teachers estimated value-added measures were lower in schools with a higher minority population as opposed to those in lower minority schools. The teachers also chose to leave school districts for what they perceived as more desirable

schools with fewer minority students (Steele et al., 2015). The researchers recommend that districts address the problem by offering incentives and creating conditions that allow teachers to improve. Although this research provides insight into which teachers are placed or left at certain schools, it fails to identify multiple factors that might also influence the decision by a teacher with a higher value-added rating to move to a school with higher perceived potential to increase their evaluation scores. Like Steele et al., research by Ingersoll (2001, 2003) points out that the least effective teachers are often placed in schools that need the strongest teachers. However, these studies fail to analyze any differences between the certification held by the new teachers. Are all urban teachers less effective than those in neighboring districts, or does certification level make a difference in the quality of teaching?

Why Do We Have a Teacher Shortage?

In response to the teacher shortage, Oklahoma modified its teacher certification policies to increase the number of teacher applicants from alternative pathways. On average, the turnover rate of first-year teachers in high-minority schools is 14.03% compared to 9.10% of first-year teachers in low-minority schools (Redding & Nguyen, 2020). Research suggests that teachers are leaving urban school districts for very specific reasons. Olsen and Anderson (2007) found that teachers become dissatisfied when administrators fail to meet their needs. Ingersoll (2004) also identified a lack of support from administrators as a primary reason why teachers leave the classroom. According to a 2005 survey of more

than one thousand California teachers, salary and school conditions were the strongest predictors of high rates of school turnover (Loeb, Darling-Hammond & Luczak, 2005).

Even teachers who have achieved success in urban settings often experience burnout due to additional duties imposed by school leadership (Maslach & Leiter, 2008). In a qualitative investigation of urban teacher retention in Los Angeles, Olsen & Anderson (2007) investigated fifteen new teachers through three in-person, semi-structured interviews over the course of a school year. Participants also completed 90-minute classroom observations. The sample selected for the study were students at Center X, a two-year teacher preparation program at UCLA that focuses on increasing racial diversity among teachers. The participants took courses during the first year and participated in student teaching. During year two, they continued their coursework while practicing full-time in a classroom setting with support from Center X's faculty. The study focused on why teachers continued employment within a given district and sought to identify any noticeable trends in responses. Several participants believed that they could not keep up the commitment to the intense work pace required of an urban teacher. Many of the teachers wanted to remain in urban education but decided to move into positions within the school district that supported students and teachers outside of a classroom setting. Several teachers joined Center X to support and motivate students in urban communities because they noticed the lack of teacher diversity when they were in K-12 school. However, these teachers felt that they could make a larger impact working in a different capacity as a principal or

district-level administrator. Three teachers noted that opportunity costs kept them in the classroom: they had already invested time and money into becoming a teacher. It is important to consider that these trends may not be generalizable to all urban districts due to the specific training provided by Center X. This group of participants received additional funding for completing the program, and many had previously participated in urban schools through mentoring or tutoring programs.

To address teacher shortages, Ashiedu and Scott-Ladd (2012) performed a qualitative study of teachers throughout Australia. The researchers found that the most important factors in teachers' decision to enter the profession were their desire to contribute to society, their enjoyment of working with children, and their love of their subject matter. Family needs and personal school experiences were less influential factors. The participants also considered pay and working conditions to be important. Teachers identified administrative responsibilities and poor discipline as negative aspects of their positions, and those who left the profession cited these aspects as major contributors to their decision to resign. Interviews revealed a strong concern about poor discipline due to leadership stress or lack of respect. The study did not differentiate between teachers in suburban and urban schools. This additional criterion might have provided insight into the differences between the different groups of teachers. However, Redding & Nguyen, using decades worth of national data from Schools and Staffing Survey, found that a pay increase of $1000 decreased a teacher's risk of leaving by 5% (2020).

Ingersoll (2001) utilized a national Schools and Staffing Survey dataset from 6733 surveys of teachers who have left the profession. With a focus on urban schools and small private schools, the study found that student discipline problems and lack of student motivation were common threads in self-reported data on dissatisfaction. Redding & Nguyen likewise study teacher dissatisfaction, finding that teachers report worsening working conditions (2020). Teachers receive less support from administrators and have more student discipline problems to face (Redding & Nguyen, 2020). Researchers have identified that student discipline has continued to be a problem for teachers over several decades. Ashiedu and Scott-Ladd (2012) proposed solutions based on improving relationships and cultures in schools. This study offers insight into other countries' issues with teachers leaving the profession and provides a foundation for future research.

Donaldson (2009) completed a longitudinal, retrospective study to inform efforts to retain urban teachers who participated in a Harvard Graduate School Masters Level Teacher Education Program (TEP). When TEP began in 1984, it attracted a variety of aspiring teachers. In 2001, TEP's focus shifted to work with urban schools. Compared to non-urban cohort members, urban teachers who participated in TEP were no more likely to exit the classroom than the non-urban teachers. However, those teachers who were in the urban-focused program were less likely to stay in teaching by means of the classroom. The TEP participants felt a lack of community support for the schools they worked at and were drawn to other roles like principals or lawyers specializing in educational issues

(Donaldson, 2009). In these new roles, they were able to address the concerns for the school sites that they noticed as a classroom teacher. It was also found that white teachers and males from urban cohorts were more likely to leave the classroom. The odds that a white female in the urban cohorts would exit were 1.81 times greater than those who were in prior cohorts; the odds for white males in the urban cohort were 4.90 times higher than the prior cohort. This leads to further questions about the role of race in teacher retention, specifically in urban districts. A major finding that supports a need for future research is that 14% of the respondents named a lack of community support as a very important factor for leaving (Donaldson, 2009). Further investigation into a lack of community support is required to understand this finding.

Teachers may not understand how to interact with the communities they serve in an effective manner and underlying racial tensions or misconceptions may keep teachers from reaching out for assistance. It is also important to note that not all urban teachers leave because of community interactions and behavior; other issues, specific to each study, are likely to play a role.

Student behavior, extreme work demands, and lack of community support are often identified as contributing factors in teachers' decisions to leave the classroom and, consequently, in the increased demand for alternative replacements. Evaluation tools, when used formatively, might remediate these issues and offer teachers support through dialogue and professional development. The MFTEM contains elements on establishing rules and procedures, establishing and maintaining effective relationships with students, and building collegiality

and professionalism. These elements allow evaluators and teachers to assess both current successes and areas for growth. With a supportive evaluator, the MFTEM may be a solution that both improves the quality of teaching and reduces teachers' desire to leave.

Teacher Preparation for Urban Schools

Freedman & Appleman (2009) conducted a five-year qualitative study on a group of beginning secondary English teachers who became part of a strong cohort. All of the teachers taught in high-poverty schools, identified as those with 50% or more of the students receiving free or reduced-price lunch, within urban school districts. The program provided new teachers with pedagogical support and a theoretical foundation focusing on social justice. The study utilized National Center for Educational Statistics definitions to define the categories of "stayers," "movers," and "leavers" for purposes of comparison to other studies. Teachers who participated in the Multicultural Urban Secondary English (MUSE) program at the University of California, Berkeley, had a 69% local retention rate, which compares favorably to Ingersoll's (2003) finding of 50% retention for teachers in high-poverty schools nationally. MUSE offers insight into potential opportunities for teacher growth and development that might decrease teacher mobility out of high-poverty urban school districts and increase teacher effectiveness through the creation of leaders focused on improving educational opportunities for students in urban communities

MUSE-based research by Freedman and Appleman (2008, 2009), Sleeter (2001), and Olsen and Anderson (2007) tout the retention of teachers who are

trained specifically to work in urban education. However, MUSE includes only a handful of graduates each year, and the studies are small and specific, which may limit generalizability. The preservice teachers commit early on to work in urban schools and self-select employment in these schools as a primary goal. Freedman and Appleman (2008, 2009) found that the MUSE cohort had a 96% retention rate, much higher than the national rate of 76%, which included all teachers regardless of the school's location or economic status. Similarly, the Center X cohort at UCLA (Olsen & Anderson, 2007) had a 98% retention rate after the first year. Both programs focused on continued support for teachers who serve English language learners. Not every teacher comes from a traditional education background geared toward urban education, but it is important to point out that by educating future teachers on social justice and the realities of working in an urban school district, these two programs achieved retention rates dramatically higher than average. If social justice is key to the success of MUSE and Center X participants, it might be an important indicator of a quality teacher in urban school settings. The MFTEM includes understanding students' interest and background as an indicator of quality teaching.

Deficit Thinking May Hinder Relationships

Some educators walk into the classroom with preconceived ideas about the students they serve (Gorski, 2016). Such ideas often come from a "deficit mindset," the (perhaps unconscious) belief that their students are somehow "lacking." For instance, they might believe that students of color are slow learners or that ELL students will not learn to read on grade level (Gorski, 2016). A deficit

mindset holds a student responsible for the challenges and inequalities they face coming from a poor, marginalized, or oppressed population (Davis & Museus, 2019). These perceptions lead to situations where teachers might not see a student's true potential, focusing on what they believe a student cannot do instead of seeing a path to growth. When a parent does not participate in school activities, teachers may see a lack of caring instead of recognizing structural barriers that keep the parent from attending (Hamlin & Li, 2020; Gorski, 2016). Professional development training in multiculturalism and diversity is known to increase teachers' success in urban schools, but there is still a disconnect between exposure and practice. To grow, students need relationships with their teachers. It is important that current research address the impact of deficit thinking because it may reduce teachers' ability to build strong relationships with students.

Pollock, Bocala, Deckman, and Dickstein-Staub (2016) found that teachers often resist district-provided diversity training. Building on a two-year analysis of a course designed to teach diversity, the researchers critically analyzed the misconceptions that participants held regarding diversity training. Participants were approximately representative of teachers in most urban settings; 57% were white women. The participants often felt as though the tasks they faced were overwhelming or impossible. The participants mischaracterized teaching for diversity and struggled to create real change in their own practices. For instance, they believed that to learn about all students' cultures meant they had to learn everything correctly the first time and interact with students' cultures in a way that never allowed for mistakes. Participants also felt that they would be judged

on how they handled racially charged comments in the classroom, so instead of asking others for help and support, they thought they had to tackle the problem on their own; thus, they usually chose not to react. However, the researchers never specifically asked for anything extraordinary. The researchers found that if the teachers viewed the task as something they could never possibly do perfectly, they just glossed over issues in diversity. This study suggests questions for future research on teachers' misconceptions about urban schools and the caution that must be used when developing curricula to aid in effective dialogue about race. The researchers interacted with the teachers through each class meeting and detailed common themes through teachers' anonymous journaling. Although the researchers discussed the uncomfortable feelings teachers experienced while talking about diversity and the loneliness they felt as they tried to complete tasks such as educating themselves about their community, teachers felt overwhelmed by the content. It would be valuable for future research to focus on teachers' perceptions about their relationships with students and to delve deeper into specific deficit mindsets. The research demonstrates a need for teachers to critically evaluate the privilege that comes along with whiteness in urban settings. The teachers and leaders who participated in the study showed a clear need for continued improvement in the area of race relationships and the discomfort prompted by this topic. When a district creates professional development on race, it is important that participants feel as though it is manageable and attainable. The teachers must know that they are not alone when

they attempt to make changes in their classroom or mindset and that group conversation may be an avenue that allows for self-exploration.

Sleeter (2001) reviewed the courses offered to white preservice teachers to counteract their stereotypes and lack of social and cultural background knowledge. Case studies and narrative research are predominantly used to reflect on preservice teachers' experiences in new settings. It is important for teachers to reflect upon their feelings about race to move forward and show growth. The research covers a variety of methods used to teach white teachers about multicultural education and racism, particularly in urban districts, because of the strong need for white teachers to understand the social and cultural norms of the students they serve. Research demonstrates the need for training in diversity, multicultural, and racism if white teachers are to be successful (Freedman & Appleman, 2007; Lowenstein, 2009, Sleeter; 2001)

Over the last few decades, several different methods have been used in college courses that train preservice teachers in topics such as multiculturalism and diversity. However, these courses generally appear to lack depth of content (Lowenstein, 2009). This includes courses that cover broad discussions on race, class, and language. They may include a focus on differences between groups and the marginalization and oppression that exist in our society (Lowenstein, 2009). Some multicultural courses also include programing on immigrant issues, parenting, disabilities, and Jewish perspectives. No one clear definition of multicultural education encompasses all the courses a prospective teacher might take. Lowenstein (2009) frames a central problem that focuses on the

conceptualization of white teachers as deficit learners on issues of diversity in multicultural education. Through a critical analysis of the literature, Lowenstein identifies inequities and assumptions that are made about white teachers. Lowenstein suggests that white teachers become disengaged with the classroom because of the assumption that they are not ready or willing to learn and identifies a need for further dialog about how learning experiences are interpreted by participants. It is important to understand how white teachers could best learn about critical race theory and multiculturalism; however, the author makes it seem as though white teachers have been the victim of a deficit mindset and have therefore been unable to learn well. This article's framing of the issue sheds light on white teachers' misconceptions about deficit mindset and their lack of understanding about the impact of whiteness. However, the article does emphasize the fact that little research exists on what really works and what ultimately makes a better teacher for a diverse population of students.

Lowenstein (2009) notes that it is difficult for a university faculty member to teach a course on multiculturalism when they themselves have had, on average, less than five years of experience with the subject. If teachers are expected to meet the needs of students, then preservice courses must adapt so that more relevant, effective, and efficient learning can occur.

Goddard and Goddard (2001) apply social cognitive theory to offer a theoretical analysis of the relationships between self-capability and collective efficacy. Efficacy is defined as the perception that one can complete a task at hand successfully. Utilizing one large urban school district's elementary schools,

the researchers surveyed over 400 teachers and their students. The students were identified as poor, racially minoritized students with below-average achievement. Roughly two-thirds of the students qualified for free or reduced-price lunch. The study was conducted utilizing fourth-grade state assessments as the proxy for academic achievement. The researchers were interested in finding out the efficacy level of teachers as individuals and groups as related to increased academic achievement. The study found that when teachers are around others who believe themselves to be successful, the teachers work more effectively with minority populations. If teachers believe they are going to be successful in the urban environment, it improves the relationship between teacher and student. When teachers had higher efficacy, a positive relationship occurred: students' positive behavior led to increased academics and improved self-efficacy. Researchers found that students developed stronger relationships with teachers who expected that achievement was possible and who utilized a variety of techniques to meet each student's needs. The MFTEM addresses this quality required of an urban teacher through the element of communicating high expectations for all students. The scoreable elements include demonstrating value and respect for low-expectancy students, asking questions, and probing incorrect answers. Low-expectancy students are those whose goals are set too high by the teacher, and the student believes they cannot reach them. It is important to understand whether differences exist in this category based on teacher certification levels so that state policy might address any deficiencies that are found.

White Teachers' Feelings While Working in Urban Schools

Patton & Jordan (2017) studied a case in which a Black woman administrator attempted to address racism in an urban school where the predominantly Black students were taught primarily by white teachers co-teaching with Black assistants. As a new administrator, she noticed racial divisions among the teaching staff and wondered whether the effects of racism were also reflected toward students in the classroom. As the staff began to address racism, the white teachers felt attacked and misrepresented. They claimed that they had not been racist and that the Black teachers were trying to look like victims. The white teachers felt as though the offered training was a waste of time, and they pushed back, threatening to report the staff discussions and training to the superintendent in hopes of suppressing the difficult dialogue. The white teachers continued to undermine the principal's attempt to address racism. Teaching about racism through whiteness is an uncomfortable process, but one that must be considered if all students are to be taught in a responsible manner. In this school, many white teachers demonstrated a method of thought that negatively impacted students' success and harmed the community. If racism exists in school, is it part of the reason why white teachers leave? More research on the impact of race and whiteness and its effect on retention is needed to identify whether any link exists.

Haviland (2008) drew from a yearlong qualitative research study using critical studies of whiteness and their implications for social inequality. The data were collected in a white-dominant eighth-grade classroom being taught by a student teacher. The student teacher reported that it was more difficult to discuss

issues of racism and white supremacy with the white students because they glossed over the issues and the status quo was often reinforced.

Teacher-Student Relationships

A meta-analysis by Roorda, Koomen, Spilt, and Oort (2011) examined 99 studies linking positive and negative relationships to engagement and student achievement. The sample was compiled of previous relevant studies that showed sufficient statistical information to calculate an effect size. Students from preschool to twelfth grade were included in the study, and measures were similar throughout. Ninety-two articles describing 99 studies between 1990 and 2011 made up the final sample. Due to the large sample, the results are generalizable to a variety of school districts. Stronger teacher-student relationships were associated with better student functioning, fewer behavior problems, and higher engagement in learning activities.

Behavior problems have been found to be one of the leading factors cited by teachers who have left urban schools (Donaldson, 2009; Ashiedu & Scott-Ladd, 2012). Positive relationships allow students to feel secure and supported, which enables teachers to engage in learning tasks. However, negative relationships lead to a sense of insecurity among students and make it difficult for them to cope with the demands of school. Students are motivated by feeling safe, respected, and valued (Hamlin, 2021; Rice-Boothe, 2009), and this occurs when the teacher shows involvement, provides structure, and supports autonomy. For teachers to be able to meet these psychological needs, they must truly understand where a student comes from and what their parents consider to be norms. It was

also found that minoritized students from households of low socioeconomic status have more to gain or lose from relationships than others, and these are the students served by most urban schools. Goddard, Tschannen-Moran, & Hoy (2001) found that, when teachers try to overcome cultural differences to build trusting relationships with students, it is more difficult to overcome economic differences than racial differences. But when teachers continue their attempts to overcome cultural differences to build trust and relationships, they are often successful in the urban school setting (Duncan-Andrade, 2007). Therefore, it is important that teachers in urban schools understand the vulnerability that exists within the teacher-student relationship so that they can monitor their actions and act with care.

Hamre & Pinata (2001) followed 179 students from kindergarten through eighth grade at a small city school to examine the extent to which teacher-student relationships predict future outcomes in the next few years of schooling. They found that when teachers' perceptions of students were positive, students were less negative and hostile throughout their school years. They also noticed that children with an insecure attachment to their teachers had elevated levels of aggressive and hostile behaviors. The researchers stress the importance of developing a strong positive relationship with students because these relationships enhance school outcomes. If a teacher fails to build strong relationships with students, teachers will have difficulty in the classroom and potentially cause future harm in other teachers' classrooms (Hamre & Pinata, 2001). A more recent study by Huang, Lewis, Cohen, Prewett, & Herman (2018) shows similar results

to those of Hamre & Pinata (2001); they found that students who had strong teacher-student relationships had fewer problems with depression, concentration, emotional regulation, and behavior engagement. These findings show a need for strong positive relationships between teachers and students to increase students' social and emotional wellbeing, particularly in urban schools. They suggest that training in this area is especially important in secondary schools.

In an investigation of teacher-student relationships in an urban school district, Huang, Lewis, Cohen, Prewett, & Herman (2018) also identify a strong relationship between the outcomes of bullying, academic outcomes, and psychosocial outcomes for students. When the teacher-student relationships were strong, the level of disruptive behaviors decreased and the level of positive behaviors increased. When teacher-student relationships were poor, they had a negative impact on student engagement (Huang et al., 2018). Although this study focuses on the impact of bullying and teacher-student relationships, it builds on the literature that shows relationships in urban schools to be a critical aspect of teacher quality and supports the idea that relationships matter in urban school settings.

How Can Teachers Build Relationships?

Teachers often express compassion for students' overall wellbeing in addition to concern about their academic performance (Adams & Forsyth, 2009). In interpersonal relationships, demonstrations of compassion build trust, so we might expect students to have more trust in teachers who exhibit compassion. However, very little research has been conducted on students' trust of

teachers. Adams & Forsyth (2009) developed a scale that assesses the trust relationship between student and teacher. Trust—the willingness to be unguarded and have a reciprocal feeling of reliability, competence, and honesty—is a necessary condition for effective school performance (Adams & Forsyth, 2009). However, transforming school structures and cultures requires risks that many teachers, parents, and community members are not willing to take. If students perceive the behavior of their teachers as open, honest, competent, benevolent, and reliable, they can believe that the teacher's future behavior will be consistent (Adams & Forsyth, 2009). This leads to a stronger relationship in which both parties trust each other and interact with honesty. The scale created by Adams & Forsyth (2009) addresses students' perspective through a reworking of the Trust Scale (Hoy & Tschannen-Moran, 1999) and the Parent Trust Scale (Forsyth, Adams, & Barnes, 2002). The scale provides a point of reference for future research, particularly for students in urban districts. It not only assesses the teacher's perception of students but also determines whether students notice any underlying tones presented by the teacher that led the student to distrust a teacher or act out based on a teacher's perception. Adams and Forsyth's (2013) study on the effect of trust in urban elementary schools highlights the importance of trust among teachers, students, and parents to improve learning outcomes. The study was completed through a series of questions utilizing a Likert scale to identify the strength of feelings of trust. Adams and Forsyth (2013) identify high relational trust as a representation of one's individual beliefs, not of the school culture. Although a higher trust score in overall school relationships is important,

the individual relationships were found to be more important. The students who were part of a school with higher trust relationships felt as though they were more in control of their learning and achieved higher scores than those who were part of a low-trust school. The students in the higher trust schools were more self-regulated, performed better socially, and had higher academic achievement.

If new teachers feel as though behavior problems and lack of respect are driving factors in their decisions to leave urban schools, then trust should be evaluated as part of future research. Is trust absent in the classrooms of teachers who leave urban schools? If trust is a factor, is it driven by underlying racial issues? As part of future research, it is important to question whether a teacher feels as though they are trusted by their students, whether they trust the community they serve, and whether trust issues are part of the reason teachers leave. A related question is, to what extent can trust be increased when racial bias is present?

Researchers identify trust and relationships as an important part of student success. The Marzano evaluation tool uses teacher relationships as a measure of teacher quality through an evaluation of a teacher's ability to maintain effective relationships, celebrating progress and communicating high expectations for all students.

Definition & Measurement of Teacher Quality

The definition of teacher quality varies greatly among researchers and practitioners. This inconsistency leads to a variety of measures of teacher quality,

ranging from student assessments to teacher observations to personal qualities and beyond (Goe, 2007).

A teacher must know what to teach and how to teach it. Teaching training has traditionally focused on two elements: subject expertise and pedagogy. A quality teacher is one who knows both what to teach and how to teach it. But Cochran-Smith (2003) points out that this traditional definition is very different from the government's technical definition of a quality teacher. The No Child Left Behind Act of 2001 defined teachers as being "high quality" if they were able to pass a series of state exams and receive full certification through the state by any means (Cochran-Smith, 2003; Keller, 2007). However, teachers who had been in the classroom prior to the act were not able to improve in quality, and teachers who passed exams were not necessarily of higher quality (Cochran-Smith, 2003). The Secretary's report from the U.S. Department of Education (2002) states that content knowledge and general verbal ability are the most important qualities of a highly qualified teacher and that teachers should utilize scripted texts to ensure consistent practices. Cochran-Smith (2003) strongly refuted the U.S. Department of Education's (2002) definition and identified a highly qualified teacher as one who selects teaching strategies that are best suited for students' learning, assuming that teaching is a complex and uncertain process instead of a script to be followed. Similarly, Darling-Hammond and Youngs (2002) countered the assumption that teacher education and certification were unrelated to effectiveness and that subject matter and knowledge were the most important attributes of effectiveness. Some critics of teacher certification argue

that not enough restrictions are placed on teacher licensing and that requirements should be adjusted to require more qualifications, as a doctor or lawyer would undergo (Ingersoll, 2004). This would allow entry into the profession to only the highest qualified persons. However, no mention was made of increasing teacher pay to the level of doctors or lawyers.

Teacher certification has also been viewed by many state-level administrators as too restrictive for prospective teachers, preventing high-quality candidates from joining the field (Ingersoll, 2004). Preservice teachers are required to take multiple assessments of knowledge to meet each state's requirement and, in many states, must also continue to attend professional development throughout the school year. Potential teachers can pursue alternative routes to education when they hold a degree in a particular subject area, because knowing what to teach is considered more important than how to teach (Ingersoll, 2004).

The National Comprehensive Center for Teacher Quality suggests a variety of ways to measure teacher quality, such as standardized achievement tests linked to student-teacher data, attributes of preparation programs, or experience (Goe, 2007). The Center's research also suggests that teacher quality might be measured using different metrics depending on the situation. For example, teacher qualifications and experience might be an effective way to measure quality in some situations, but having the ability to motivate students who are at-risk may measure quality in other environments (Goe, 2007). Indicators of quality relevant to hiring a teacher may differ from those used when considering tenure (Goe,

2007). However, researchers agree that no one definition of teacher quality exists; instead, definitions are manipulated to suit the occasion or purpose for which they are being used (Goe, 2007). Goe (2007) describes potential inputs for teacher quality as education, certification, credentials, experience, race, gender, beliefs, and self-efficacy. Race and gender allow researchers to specifically identify any impact on student learning caused by these two factors that teachers cannot change. Literature in this area is lacking. Goe (2007) defines teacher practices as evidence of teacher quality both in and outside the classroom. Both the inputs and the processes create a framework for teacher quality that is then generally measured by student achievement data.

Because of its ease of access, student achievement data has been used to link teacher performance to students. This is becoming more prevalent as districts more frequently maintain integrated data systems that hold long-term data on students. Boyd, Lankford, Loeb, and Wyckoff (2013) discuss the inaccuracy and unreliability of systems that use of student achievement as an outcome measure: Teachers cannot control many factors outside of the classroom, yet these factors can impact student learning. As referenced in Chapter 1, four broad phenomena must exist for learning to occur. Students must actively participate in learning, and they must have a supportive environment, opportunities to learn, and good teaching (Fenstermacher & Richardson, 2005). Economic factors affect student performance on achievement tests; it is well established that differences in scores exist among socioeconomic groups. This should be considered when calculating value-added scores, because a student's socioeconomic status is a factor outside

the teacher's control and should not count as an indicator of teaching quality (Boyd, Lankford, Loeb & Wyckoff, 2013). Additionally, students at the secondary level move from one teacher to the next throughout the day, so researchers may be unable to account for the influences of science teachers on math and reading assessment scores. The No Child Left Behind Act of 2001 and the Every Student Succeeds Act (2015) require districts to track and report numerical data for accreditation purposes. Due to the easy access to such data, many researchers have utilized student achievement data as the primary measure of teacher effectiveness.

Instead of static snapshots of student performance on standardized tests, practitioners such as principals or teacher leaders may have a better perspective on teacher quality through regular observations of novice teachers. Hollins (2011) defines quality teaching as a complex system of levels centered on a strong understanding of essential knowledge, skills, and understanding. Quality teachers must understand the learning process combined with new theoretical perspectives on learning frameworks (Hollins, 2011). They must have a vision and purpose for defining outcomes and developing classroom assessments at the formative and summative levels. Hollins (2011) also believes that teachers of quality understand the value of observing and collaborating with other teachers and practicing the skills required to lead a classroom effectively. Peer observations and deep engagement in communities of practice provide insight into what other teachers implement within the classroom. Teachers of quality are willing to collaborate and grow professionally with each other (Certo & Fox, 2002, Killion, 2015).

Some researchers define quality teaching based on preservice preparation programs. Von Hipples and Bellows (2018) compared teacher preparation programs and value-added reports in six states to evaluate the quality of teacher preparation programs. This research used teacher value-added evaluation reports to determine whether a teacher was "of quality," then worked backward to identify whether the teacher's specific school preparation program was a factor in the quality differences. They found few differences between programs producing significantly stronger quality teachers and programs that produced weaker teachers.

Researchers often define quality teachers as those who have higher value-added scores based on student growth (Boyd et al., 2013; Joyce et al., 2018). Student growth is often determined by a series of longitudinal achievement tests administered at the state or national level. However, these methods are often flawed due to test bias, which skews the data. Test bias can be introduced by language that is common to one group of test takers but unfamiliar to others. For instance, a student from an urban school was unable to understand a reading passage that mentioned a "gazebo" because gazebos were not part of his lived experience. He thought a "gazebo" might be a type of lizard, which affected his comprehension of the reading. It can also stem from readings describing situations that may be native for middle-class communities but not accessible to students from lower socioeconomic groups. Another problem with tests is that they may not align well to the academic standards being taught in the classroom, making it difficult for students to demonstrate mastery (Boyd et al., 2013).

Joyce, Gitomer, and Iaconangelo (2018) found that a more accurate measure of teacher quality could be found through quizzes and classroom assignments, which shed light on the intellectual demand required by teachers of students. “Intellectual demands” encompass how students respond to challenging tasks and how their assignments affect student outcomes (Joyce et al., 2018). When teachers ask open-ended questions or questions with more than one correct answer, students struggle due to the increased academic rigor or intellectual demand. Through an analysis of classroom artifacts, such as activities, quizzes, and assessments, evaluators can identify teachers who lack quality throughout the school year. This allows evaluators to assign training and professional development to teachers who produce less rigorous artifacts of lower quality throughout the year. Teachers then could become higher quality teachers in a proactive, professional manner. In many cases, only state-based assessments are used to determine student growth; however, Joyce et al. (2018) identified a way to look at the assessment or artifact data preventatively to promote quality teaching prior to the final evaluation of a teacher’s performance in the classroom. Through ongoing analysis of artifacts in the classroom, Joyce et al. (2018) recommends measuring teacher quality frequently and recommending training consistently to continuously improve one’s practice.

Additional literature suggests that students also provide a perspective on the definition of a quality teacher. Sabine (1971) identified themes of challenging, caring, and listening as student-defined indicators of teacher quality. Cochran-Smith (2003) also found that caring relationships formed with students indicate a

quality teacher because students want to learn from someone who cares for them; and quality teachers have strong communication and teaching skills and foster interest in the course. Jones and Henry (2019) found that students identified quality time spent on content as an area of general teacher improvement that affects students' perceptions of teacher quality. However, caring *about* students instead of caring *for* them may result in no benefit for the students because these teachers focus on issues only instead of finding solutions to them (Shevalier & McKenzie, 2012). Caring *for* someone occurs when teachers and students have face-to-face relationships, experience the issues together, see the consequences, and reciprocate the cycle of caring (Shevalier & McKenzie, 2012). When teachers build caring relationships with culturally different students, they respond to them in ways that reinforce positive connections, which are imperative to quality teaching.

Common themes throughout the literature focus on professional development and student growth as factors in identifying quality teaching. Evidence of these two areas might be found through classroom evaluations if the evaluator has been trained to provide accurate analysis. Marzano and Toth (2017) believe that teacher quality can be assessed through the MFTEM if it is used appropriately. The evaluation tool reports on a variety of input measures, including student engagement and scaffolded instruction, while also utilizing very specific target measures, including growth and quality of in-classroom practices (Marzano & Toth, 2017).

No clear definition of a quality teacher can be determined by any one policy-making body, legislator, or researcher. Each definition has flaws that can create inconsistency and confusion. This is similarly the case when trying to measure quality. Often current research does not separate teachers by certification type before identifying factors of growth. Generally, all teachers are aggregated to the same category or group, such as urban or suburban teachers (Duckworth, Quin & Seligman, 2009; Sachs, 2004). It is important to identify any differences between certification type and perceived measures of quality so that knowledge about professional personnel can guide policy and leadership practices that better ensure a prepared and effective teaching force, ensuring that all students have access to quality teachers. For the purposes of this study, we will utilize Marzano's (2007) definition of teacher quality that accounts for many aspects of the aforementioned definitions into one evaluation tool.

Quality and Effectiveness

In the published literature, the terms "quality" and "effectiveness" often are used interchangeably; however, the terms appear to have slightly different meanings for different researchers. Quality may be measured by inputs, and effectiveness may be measured by outputs. Ronfeldt, Lobe, and Wyckoff (2013) believe that teacher effectiveness is something that individuals bring with them to the classroom; it is not taught but is part of a total package that someone inherently possesses or does not possess. Teacher effectiveness is also defined as the value a teacher adds to student achievement according to a scale. However, this definition may not account for students who have learning disabilities or are

labeled at-risk. These students may show growth each year, but this growth may not be enough for the teacher to earn the label of quality teacher.

Value-added measures, utilizing state or national exams, fail to evaluate teachers beyond the math and reading classroom. Many studies utilize only student achievement on state or national reading tests and implicate a narrow band of associated teachers. The measurement of teacher quality should have a more holistic approach to identifying teachers who excel in all areas of education, not just math and reading.

In business, "effectiveness" is defined as producing the intended or expected result or being successful in producing a desired or intended result (Van Cauwenbergh, 2010). "Quality," however, is measured by how well a product aligns to an industry standard, usually based on customer satisfaction, and includes continuous improvement (Van Cauwenbergh, 2010). Through observations of teacher practices and interactions with students and peers, we might get a clear picture of what it takes to be a quality teacher. Researchers argue that teacher quality should be measured by the inputs that teachers align to state standards like lesson plans and assessments, yet current research does not address this issue (Boyd, Lankford, Loeb, & Wyckoff, 2013).

Because quality is measured by inputs and effectiveness is generally measured by outputs, a teacher's quality should be evaluated before looking at measures of effectiveness. Outputs could be viewed as student learning within the classroom including exit tickets, productive struggle, or dialog instead of regulated state or national assessments that are normed to populations that may

not be represented in urban school settings. Exit tickets are short questions that students answer on their way out of the classroom to summarize learning or indicate areas they did not fully understand during the class period. Teaching quality could be measured by the inputs instructors possess in knowledge, dispositions, and credentialing to be utilized in creating an environment for learning.

Effectiveness Through Standardized Assessments

Effectiveness has most often been measured through student assessment that aligns with state standards (Boyd et al., 2013). Student growth or annual yearly progress is measured through an assessment tool, and a measure is put in place to identify how much influence a given teacher may have had on the student's academic growth. The teacher is then ranked as effective or ineffective, which is an outcome of quality teaching. Historically, this method has been considered a valid way to identify quality teaching, even though standardized achievement tests were not designed to measure teacher quality (Goe, 2007).

Goe (2007) also identifies the difficulty of sorting out the differences between the effect a teacher has on standardized achievement tests scores and the effects that peers, textbooks, and other factors might have. This is the case even when value-added assessment scores ostensibly account for these co-variants within the metric (Goe, 2007). Wilson and Floden (2003) found enough inconsistent findings and analysis about teacher effectiveness that it was difficult to find evidence strongly supporting the findings in each study. For instance, when synthesizing the research of 12 studies, Wilson and Floden (2003) found

that the relationship between teacher practices and student achievement were inconsistent, rendering overall findings invalid due to questionable research design or methods. They also found that many of the results were not statistically significant even when some positive correlations were noted. However, Kini & Podolsky (2016) found that as teachers gained experience in the same subject area, students learned more on standardized tests and had better school attendance.

If previous research studies had been able to apply recent technologies and new ways to link teacher data to student data, they might have been able to produce more accurate data (Wilson & Floden, 2003). Boyd et al. (2013) identifies errors in student testing that result in an error in teacher evaluation based on an analysis of math and English language arts (ELA) test scores in New York City. The researchers found test measurement error to be at least twice as large as the vendor reported to the state. This error may have been due to missing data, non-parallel assessments, frequency of testing, or other unknown factors (Boyd et al., 2013). It was found that only 20% of the variance in math score gains was due to prior achievement and that this percentage was higher than gains in ELA (Boyd et al., 2013). Noell, Burns, & Gansle (2019) developed a value-added assessment of teachers in Louisiana. During the development, researchers noticed that teachers were resistant due to the top-down mandates for the new model. They stress the importance of continuously questioning the tools and measures used in value-added assessments because the systems are complex and challenges exist when developing measures of quality and effectiveness (Noell et

al., 2019). It is important to understand the inaccuracies associated with current methods of assessing teacher quality through an outcome that has many issues and might not be considered the most useful or accurate method.

In Oklahoma, schools are now required to assess less frequently than they were before 2017. Secondary mathematics is tested only through a national assessment that might not parallel the state standards or school district expectations. Therefore, an error in analysis would inherently exist if this assessment data were to be used as a measure of teacher quality. The school district selected for analysis in this study has chosen a different student assessment three years in a row for the secondary level. They have moved between the ACT and SAT as an assessment of student learning. This provides no solid history of evaluation at this time, as students who took an assessment in the 2019-2020 school year were assessed on a pre-assessment that utilized different measures. For this study, a student assessment will not be utilized as the output measure to identify teacher quality due to its apparent ineffectiveness and the unavailability of sufficient longitudinal data to make valid correlations between teacher certification type, quality, and student performance.

Performance of Emergency-certified Teachers

When using VAM, Darling-Hammond (2010) found that teachers with the largest impact on student achievement were those who had earned a standard teaching license in the field, had taught for two years, and had passed certification tests. Boyd, Lankford, Loeb, Rockoff, & Wyckoff (2008) found that among teachers in New York City, uncertified teachers had the weakest value added

when they investigated fourth- and fifth-grade math teachers. Kane, Rockoff, & Staiger (2008) also note that although little or no differences in average teacher effectiveness between certification levels, the difference that does exist is the equivalent of teaching for eight years in that school district. When teachers are emergency-certified in the state of Oklahoma, they are not required to pass any licensing assessments and have had no field experience, so they would not have the largest impact on student achievement. Using VAM, the Louisiana Board of Regents supports similar findings; emergency certification was a significant negative predictor of student achievement compared to teachers who held a standard certificate (2008). However, Goldhaber (2019) found that only minor differences exist between those who attend a teacher preparation program and those who do not. He also pointed out that data is often deficient, with sample sizes that are too small or that do not account for all graduates of teacher preparation programs.

What Makes a Good Evaluation?

A good evaluation gives teachers satisfaction in their work. Good evaluations increase job satisfaction when they motivate teachers through intrinsic values (Ford, Urick & Wilson, 2018). Shen, Leslie, Spybrook, & Ma (2012) utilized school staffing surveys to identify which variables, principal background, and school processes had the most influence on teachers' job satisfaction. The study identified intrinsic motivators including competence, autonomy, and relationships as significant predictors of job satisfaction (Shen et al., 2012). When evaluations focus on formative measures that allow teachers the opportunity to

develop learning targets and grow as educators in meaningful ways, they become a good evaluation (Reinhorn, Johnson, & Simon, 2017). Ford et al. (2018) identified a good evaluation as having frequent support of the learner through autonomous actions. The MFTEM does this through its professional learning focus (PLF). Teachers participate in a series of trainings from Marzano to develop individual PLFs that include actionable goals. Teachers share these goals with the evaluator through the online portal, and the PLF becomes part of the evaluation process. The evaluator and teacher work together to develop a plan of action, which can be as involved as the teacher wants it to be. The PLF allows the teacher to determine how they want to receive the additional professional development. They may prefer online videos that are available through Marzano, in-person professional development opportunities, or self-study through a review of research-based practices published in peer-reviewed journals.

Ford et al. (2018) identified a second trait of good evaluations: high standards with clear expectations. The MFTEM meets this criterion of a good evaluation by using a rubric that identifies what an evaluator should be looking for in a classroom to award points. The teachers are provided the rubrics in advance during district training sessions, and they are posted on the web portal so that all teachers have access to the information before beginning the official evaluation. When teachers feel an evaluation assesses them using high standards with clear expectations, they are more likely to experience an increase in job satisfaction (Lavigne, 2014). An additional element includes an assessment of teaching using more than one measure (Ford et al., 2018). Teachers expressed the

belief that being evaluated only on student assessment scores was unfair; moreover, it reduced their sense that the measure was effective and important. To remedy this concern, the MFTEM blends several areas, including classroom strategies, planning, and collegiality into its evaluation tool.

The fourth characteristic focuses on support for the purpose of improvement. To improve classroom performance, teachers must have proper support and professional development throughout the entire evaluation process (Ford et al., 2018). This includes constructive feedback and innovative approaches that allow teachers to have autonomy and feel competent throughout the process. The MFTEM has the teacher and evaluator sit down together for a pre- and post-conference. During the pre-conference, the evaluator asks the teacher what will be taught, how they developed the lesson, and what measures they will use to ensure that student learning occurs. The teacher has the time to talk about what they are trying to accomplish professionally and what they want students to learn. The post-conference allows teachers to develop an action plan in collaboration with the evaluator to determine areas where they might benefit from additional professional development and an appropriate timeline for the teacher to obtain additional support. Previous research outlines specific characteristics of good evaluation, including measures that increase teacher job satisfaction through psychological needs that can only be met through intrinsic motivation (Ford et al. 2018). The MFTEM utilizes the characteristics identified by previous research, making it a good evaluation tool.

The Evaluator

It is important to consider the role of principals as users of the MFTEM in the evaluation process. The role of principal has expanded; in addition to being a manager, employer and supervisor, the principal is responsible for student learning through instructional leadership (Kraft & Gilmour, 2016). Instructional leadership includes staff and curriculum development, analysis of student assessment, and individual teacher feedback on evaluations.

Until about 20 years ago, principals typically completed a one-time evaluation of a teacher using checklists instead of rubric-based assessments (Kraft & Gilmour, 2016). Before new evaluation models became prevalent, principals spent less than 6% of their time observing teachers, evaluating teachers, and providing coaching (Horng, Klasik, & Loeb, 2010). The new models for teacher observations use rubrics and require principals to provide feedback when scores are lower and to conduct immediate post-observation critiques that are usually in-person meetings (Kraft & Gilmour, 2016). This process takes significantly more time, with 25% of a principal's time being used for observation and feedback (Halverson, Kelly, & Kimball, 2004). This has resulted in a lack of critical feedback because principals lacked evaluation skills and required additional support (Halverson et al., 2004) and because it is more time-consuming to provide positive reinforcement rather than critical commentary on teacher improvement (Kraft & Gilmour, 2016).

During a pilot study in Chicago, researchers found that a new evaluation system, based on rubrics and feedback, produced significant improvements in students' reading when principals had comprehensive training (Steinberg &

Sartain, 2015). The research found that the nature of implementation was critically important to the potential impact of the new evaluation model (Steinberg & Sartain, 2015). However, in the second year of implementation, the same district was under new administration that favored performance and data over a rubric that attempted to improve the quality of teaching and instruction. The lack of training led to poor use of the evaluation tool in the second year, and no significant correlation was found between the teacher evaluation scores and student performance (Steinberg & Sartain, 2015).

Kraft & Gilmour (2016), using a case study of 24 principals, found that many principals were learning about the evaluation process on the job. The principals reported that they were sometimes asked to work with 12 or more teachers and felt that they could never give each person enough individualized attention to make a real difference in teachers' practice (Kraft & Gilmour, 2016). These principals also expressed concern that operations responsibilities, like emails and phone calls, decreased their time available for conversations with teachers. The principals asked for more training on how to manage time and develop the skills required to provide quality feedback.

Marzano Focused Teacher Evaluation Model

As teachers move from one school to the next, previous evaluation scores may be lost or interpreted incorrectly due to inaccuracies associated with the evaluation tools (Marzano & Toth, 2017). Before data were kept in online warehouses like iObservation, teacher evaluations were completed on paper each year. The evaluators had areas in which to score teachers, but evaluations may not

have been fair or effective due to the ambiguous wording of the evaluation (Marzano & Toth, 2017). Marzano created the MFTEM to streamline the evaluation process and create a manageable, efficient, and equitable evaluation system for all districts. The mission of the MFTEM "is to empower schools and districts to transform core instruction and leadership practices resulting in rapid gains in student learning" and focus teacher attention on specific behaviors correlated to student achievement (Marzano & Toth, 2017).

The model also includes a system that allows teachers to improve skills over the school year through a growth mindset model. Marzano (2012) suggests that a teacher evaluation system that focuses on learning and growth differs from one that attempts to measure teacher competence. Marzano and Toth (2017) believe that the evaluation tool must be fair, objective, and supportive of instructional growth if it is to improve teacher quality and performance. They believe that models must prioritize standards and allow for accurate scoring with detailed scales. Marzano (2012) identified three characteristics of teacher evaluation systems that would help teachers improve: the system must be comprehensive and specific; it must include a development scale; and it must acknowledge and reward teacher growth. Marzano and Toth (2017) believe that their model does all of this in a streamlined, accurate way through their online evaluation tool, iObservation.

Student evidence is the key factor in learning, not just a standardized assessment like other value-added models used to evaluate teachers. Through the evaluation process, teachers create clear goals and measurable action steps to

grow and improve practices. Marzano and Toth (2017) found that when teachers are given clear goals and practice in specific areas, they were able to improve practices. Research has shown that the observation score was the second largest predictor of a value-added model (VAM), and the elements in the MFTEM have a small positive correlation to student learning gains (Marzano & Toth, 2017).

Marzano selected iObservation as the only platform to be used as an implementation tool for teacher evaluation. iObservation is a data management tool that collects and reports longitudinal data on every teacher within a school district. It is an online platform on which teachers can interact with observation data and learn from supplemental resources, such as rubrics, scales, teacher and student evidence, and videos of model teaching. Evaluators can submit scores and feedback for teachers online through the iObservation portal. This allows districts to track and assess teacher performance remotely and keeps the observations secure.

Conclusion

Previous research strongly suggests that schools with the highest needs employ teachers who receive lower value-added scores (Freedman & Appleman, 2009; Ingersoll, 2004, 2002; Jacob, 2007). These teachers often are alternatively or emergency-certified and have no prior knowledge of the pedogological and multicultural coursework completed in most undergraduate programs. Even if they had completed a course in this area, research has shown that it may have had little effect on the decisions of a white classroom teacher, because often those teachers teach with a deficit lens (Leland & Harste, 2005, Lowenstein, 2009).

Research has demonstrated an increase in teacher education focused on understanding how race plays a role in institutions of learning (Ladson-Billings, 1999; Sleeter, 2017). Studies focused on students of color indicate that white teachers often act in a racially charged manner that causes discord in the classroom, which can lead to poor student behavior choices, causing stress for both teachers and students. Research has shown that when administrators attempt to address race issues within a white-dominated space, white teachers feel attacked, making them less likely to see other viewpoints; as a result, communities continue to suffer (Leland & Harste, 2005; Patton & Jordan, 2017).

Additional research focused on teacher quality by certification type may provide a clearer picture regarding the effectiveness of current policy and practice in Oklahoma. Research of this kind may also lend insight into the viability of instructional leadership of principals charged with evaluating teachers. Researchers present mixed findings when attempting to identify a correlation between teacher quality measures and student achievement. Parents and educators want to know that quality teachers are in every classroom, yet research and evaluative practices may not identify quality teachers in an effective and timely way. Instead, I suggest research should utilize the predictive factors of quality teaching through regular observations to determine whether certification type matters when determining a quality education. Future research may investigate the relationship between the low rate of teacher retention in urban schools and the subtle and unspoken racism that appears in traditional and non-traditional teacher preparation programs.

Center X and MUSE are examples of programs that fostered a positive, supportive environment to create successful teachers and an increased retention rate in urban schools Based at UCLA and UC Berkeley, these programs benefited from being housed at two of the most impressive universities in the nation, but the teachers participating in these programs lived the same daily experience as thousands of other teachers who work in urban settings. These programs therefore hold promise as examples for similar programs in other urban school districts.

The literature demonstrates a need for strong, trusting teacher-student relationships so that all students can flourish in the classroom. With a strong relationship comes increased performance, on-task behavior, and a desire to learn independently, and these are all manifestations of teacher quality (Goddard, Tschannen-Moran, & Hoy, 2001). Without positive and supportive relationships, students in urban schools will test their teachers' tolerance of off-task or disruptive behavior, and they will therefore struggle to make the same gains as other students. One tool that can aid in the evaluation of relationships with students is the MFTEM. The tool contains components that evaluate how a teacher interacts with students and how they provide positive reassurance in the classroom. Shevalier & McKenzie (2012) identify the ability to care for students as a culturally responsive practice that is critical for a quality teacher in an urban school. Although this tool may not solve all the issues surrounding race in urban schools, it can provide foundational information on teachers' ability and willingness to support students.

Racial inequities, stereotypes, bias, and other forms of deficit thinking may prevent teachers from being able to create strong relationships with students in urban schools, but it is important to recognize the problem and find solutions. Otherwise, these failed relationships may become the catalyst for the list of reasons new teachers leave urban school districts. Future research may also evaluate whether teachers who remain committed to urban classrooms show a positive relationship with students and therefore experience fewer behavior problems and less desire to leave. Teachers who stay in urban schools may show a more positive relationship with students and a decrease in racially driven practices. If this is the case, teachers will need to work together to create meaningful experiences as they proactively identify racial issues within the school that are causing conflict and stress. Without providing this type of pertinent teaching in preservice programs, white teachers will continue to approach teaching with grandiose ideas and then leave the profession prematurely deflated and burnt out.

Teacher quality does not have one specific definition that meets every evaluation's needs. Teacher quality should be measured and evaluated based on teacher performance including multiple student measures throughout an evaluation.

For many years, research has relied heavily on standardized, required assessments, but this may not be the most effective way to assess the quality of all educators. Instead, research should consider utilizing the teacher evaluation with

regular observation and feedback to determine whether certification type matters when making determinations about quality in teaching.

Emergency-certified teachers generally have reduced student outcomes when measured through value-added methods. The methodology behind the use of VAM indicates that flaws in the measure may not accurately assess teacher quality. However, the research fails to consider the teacher evaluation tool as a way to understand how certification type might impact student learning. Teacher evaluations are completed by a practitioner in the field and include some variables that are directly controlled by teacher. The process, when used with fidelity, lends itself to growth mindset and encourages teachers to grow in the profession. This study attempts to investigate the relationship between teacher evaluation scores and certification status to fill this gap in the literature.

Chapter Three: Methods

Introduction

Most research on teacher quality relies heavily on student data as a primary indicator of quality teaching. Yet practitioners often feel that student data fails to provide an accurate depiction of quality teaching. In an attempt to explore alternate measures of teacher quality, the teacher evaluation should be considered as a tool to measure teacher quality and effectiveness. Through data-proven measures developed by practitioners, this study attempts to investigate whether a relationship between teacher evaluation scores and certification status exists.

> H^1: A difference exists in teacher quality between teachers who hold traditional teacher certificates and those who hold alternative or emergency certificates.
>
> H^2: A difference in teacher quality exists between new teachers who hold a traditional teacher certificate and those who hold an emergency or alternative teaching certificate.

In this chapter I describe the research design and why it was selected to test my hypotheses. Next, I describe the participants and the characteristics of school district. Afterward, I describe my data analysis techniques and limits of the study.

Research Design

A correlational research design allows this study to investigate the relationship between teacher evaluation scores and certification status. The study

attempts to determine whether any specific categories used to evaluate teacher quality reveal statistically significant differences based on a teacher's certification level. Existing district data provides a foundation for research, eliminating potential concerns about external validity. The Marzano Focused Teacher Evaluation Model (MFTEM) has been validated by thousands of researchers, including the U.S. Department of Education (Marzano, 2013). External validity may be compromised by the evaluator's situational factors, which can contribute to a teacher's score. However, in the sample, evaluators use MFTEM on several occasions to find successful teaching practices, and scores are averaged over time. Therefore, even if situational factors affect a score, they may not be reflected in the final evaluation. I requested data from the school district using the open records request procedures. The school districts are required to report teacher evaluation scores to the state of Oklahoma each year as an accountability measure. The state of Oklahoma has been unable to link teacher certification directly to a teacher's evaluation score which is why, for this research, I requested information directly from the school district.

Teacher Certification in Oklahoma

Oklahoma has relaxed its standards for alternatively certified teachers to help reduce vacancies across the state. It is imperative that school districts understand how these new regulations might affect teacher quality. If alternatively certified teachers have lower evaluation scores than traditionally certified teachers, it is important to develop solutions to the problem so that all students have access to quality teachers. The pragmatic worldview generally meets the

needs of a mixed-methods approach. This study will only utilize a quantitative process as part of the exploratory process of the initial stages of research. Further research may be proposed for future qualitative studies if opportunities develop after an analysis of the quantitative data. The pragmatic worldview offers flexibility in methods and procedures that are not seen in other worldviews (Creswell, 2017). It allows the researcher to identify current problems and develop solutions within the means they have available, provided that they fit the situation. Through an evaluation of current policy, this study is intended to identify predictive factors of quality teaching practices between alternatively and traditionally certified teachers and provide insight into issues that may inform policy changes or alternatives.

Data Sources

This study uses both extant state teacher evaluations and teacher administrative data provided by the school district. The evaluation data includes MFTEM scores acquired from a large urban school district in Oklahoma. Probationary teachers employed less than three years and career teachers participate in the MFTEM. Evaluation data are regularly collected as part of the teacher evaluation system in Oklahoma and are required by all districts as a measure of teacher performance and quality. The data set is stored online through iObservation, and final evaluation scores are submitted to the Oklahoma State Department of Education.

Researchers regularly utilize additional quantitative measures, including national student data collected through the SAT and ACT tests, as a means of

identifying quality teaching at the secondary level. However, researchers have found significant errors in the analysis of student growth through student achievement data, leading me to choose a different direction for this study (Boyd, Lankford, Loeb, & Wyckoff, 2013). The state of Oklahoma has also recently changed and re-evaluated testing requirements throughout all grade levels, and some teachers argue that the tests are not aligned to district and state subject area standards. Due to the potential errors associated with value-added models, this study will not include these as measures of teacher quality.

Data

The Oklahoma school district selected for this study exemplifies a struggle seen across the state. An increasing number of emergency-certified teachers are being hired to fill vacancies for positions that require high-quality certified teachers who know how to work with students in urban school settings (Sutcher, Darling-Hammond, & Carver-Thomas, 2016). Emergency certifications are increasing not only in this school district but across the whole state (OSDE, 2018). To compare the quality of alternative certification routes to traditionally certified teachers, my study needed a large pool of educators who would fall into these two categories. The selected district meets this criterion. The second criterion for this study was whether the school district had implemented the MFTEM, because research strongly supports the Marzano model as a way to evaluate and improve teacher effectiveness. The selected school district uses MFTEM, providing continuous professional development for teachers and

evaluators so that all participants understand the research and procedures required to implement the evaluation process with fidelity.

The data set includes all those who teach in an urban Oklahoma school district. For this study, "urban school district" is defined as a school district inside a city with a population over 250,000. Most of the district's students qualify for free or reduced-price lunch, meaning that they have been identified as needing Supplementary Education Services (SES). The urban district selected as the sample currently identifies 50% of the total student population as Hispanic and 31.6 % as English language Learners. Study participants must hold or be in the process of acquiring a teaching certificate. Only one urban district in Oklahoma will be utilized in this study, and all respondents are employed within that district.

The sample is representative of the population, and a variety of grade levels and multiple school sites have been included as participants in the data set. To protect anonymity, all respondents' information will be free of individual identifiers such as name, school, and teacher ID number. Participant data may be shared as aggregated results if significant findings are made.

Students in the selected district are served by educators who may not have received formal training prior to entering the classroom or may have come from suburban or rural settings where students' intersectional identities and needs are different and where schools' priorities and systems function differently. These teachers have the desire to educate urban youth but may not understand how to meet the needs of the students they serve. Students in urban areas have different needs from those who live beyond the city's boundaries. Urban students may

come from backgrounds or cultures with which these teachers are unfamiliar. Differences in backgrounds have the potential to lead to misalignment of pedagogical strategies, ineffective classroom management techniques, and potential silencing of cultural voices and diversity in the curriculum, all of which can harm teacher-student relationships. Teachers who build caring relationships with students found that students were more engaged and showed improvement in academic skills and behavior (DiBara, 2015). This is especially important in urban schools where most students are labeled as at risk.

The number of participants included in the sample for this study relied upon the number of completed teacher evaluations. During the 2018-19 school year, the data contained 1285 participants in the sample set, with 882 new teachers and 403 career teachers. Cohen recommends that with an α=.05 and two groups, a sample of 393 would be required for small effect size and a sample of 26 for a large effect size (1992). The school district boasts over 4,600 administrators and teachers throughout the district. The 1285 participants are representative of teachers across the district.

Data acquired through the school district's research department has been cleaned and sorted through SPSS, a statistical software program. Having a teaching certificate and being in the classroom were required of all participants, so administrators and non-certified staff were removed from the sample. Sociodemographic data such as gender, certification type, and degree level were acquired along with the evaluation data set received by the school district. Table 1 displays the variable characteristics included in the final sample.

Table 1. Variable characteristics included in the final sample.

Variable	N	Mean	Std. Deviation	Min	Max
Total Evaluation Score	1285	3.578	0.490	1.4	5
Critical Content	1209	3.570	0.720	1	5
Rules & Procedures	1045	3.756	0.684	1	5
Policies & Procedures	1156	3.693	0.692	1	5
Standard	1285	0.621	0.485	0	1
Emergency	1285	0.226	0.418	0	1
Alternate	1285	0.153	0.360	0	1
Gender	1285	0.732	0.443	0	1
Black	1285	0.216	0.412	0	1
White	1285	0.628	0.483	0	1
Hispanic Latino/a	1285	0.071	0.257	0	1
Other	1285	0.085	0.279	0	1
Degree Level	1285	0.683	0.465	0	1
High School	1285	0.258	0.438	0	1
Middle School	1285	0.143	0.350	0	1
Elementary	1285	0.598	0.490	0	1
Career Teacher	1285	0.314	0.464	0	1
Valid N (listwise)	925				

Std. Deviation uses N rather than N-1 in denominators.

Teach for America Corps members were identified in the data. These teachers have specific, required training that they frequently attend, and they are

provided additional supports throughout the school year that are not available to all alternatively certified teachers. Some evidence points towards superior performance by several intensive alternative certification programs, similar to characteristics in Teach for America, and these special cases require a recategorization of certification types to account for the direct, specific, and mandatory training they require. However, in the data set, this group was recategorized into the alternative teaching category because, at n=10, the sample was not large enough on its own to have predictive power.

The sample was a convenience sample due to the researcher's familiarity with the school district and its proximity as the closest urban school district. Ravid (2019) cautions that care should be used when attempting to generalize the results to a larger population if convenience samples are used. In this study, time constraints keep research from going beyond the selected district. If findings are significant, it would be recommended that further research be conducted in other school districts in an attempt to generalize the findings throughout Oklahoma. For this study, it would not have been possible to complete random samples where every member of the population—all Oklahoma teachers—had an equal and independent chance of being selected (Ravid, 2019). Oklahoma does not utilize one specific model for teacher evaluation, instead allowing each school district to select a program that meets its needs. If teachers had been selected from multiple districts, additional comparisons and adjustments would have been necessary to account for the differences in measures.

Teacher Evaluation, the Dependent Variable

The dependent variables in this study include the teachers' final evaluation score for the 2018-19 academic year and the sub-scores from each of the 23 indicators of the MFTEM. Of the evaluation models approved by the state of Oklahoma, the MFTEM was selected for use in this study because it was the tool selected by the sample school district to meet the state's requirement for teacher evaluation and Every Student Succeeds Act (ESSA) in 38 Oklahoma school districts. Alternative teacher evaluation tools have had observation scores that produced a small to moderate correlation between teacher evaluation scores and value-added measures (Basileo & Toth, 2019). Teachers felt hopeful about improving instruction when they received observations utilizing the MFTEM; however, they did not make meaningful changes based on value-added measures' scores (Pressley, Roehrig, & Turner, 2018). Although findings by Basileo & Toth (2019) identified a lack of large significant findings for other observation methods such as the Framework for Teaching, the Mathematical Quality of Instruction, and the Protocol for Language Arts Teaching Observation, they were able to isolate the MFTEM to find small, positive, and statistically significant relationships between observation scores and value-added measures. The MFTEM has been shown to be a positive predictor of teacher quality in the state of Florida, and these findings may be generalizable to other states, including Oklahoma (Basileo & Toth, 2019).

The MFTEM utilizes an interval scale to evaluate effectiveness. An interval scale has equal intervals between the points (Ravid, 2019). Evaluators can give teachers a score of zero, one, two, three, or four on each of the 23 items;

partial scores (such as “3.5”) are not used. A score of zero on one of the 23 items means that the teacher has not been using that specific element. The maximum score of four indicates that the teacher utilizes innovative practices that impact student learning. The computer platform, iObservation, averages informal and formal evaluation scores together for a final evaluation score, which can result in a score that includes a decimal. For the purposes of this study, the summative evaluation score will be utilized to identify any differences in teacher quality.

The Evaluator

One concern identified in previous research is principals’ inability to properly understand effective practices, develop collaborative learning groups, or understand how the evaluation tool is part of school improvement practices (Connally, & Tooley, 2016; Darling-Hammond et al., 2012; Hallinger & Heck, 2014). To mitigate this concern, the state of Oklahoma mandates that all evaluators receive direct and accurate instruction in the implementation of teacher evaluation systems regardless of which option the district selects. The district selected for this study mandates that all evaluators hold full certificates as principals and are in the position of principal or assistant principal before evaluations begin. Principal interns are not able to conduct official teacher evaluations; however, they do go through the training process. The evaluators must also pass a written and inter-rater reliability exam before they can evaluate teachers.

The MFTEM tool has a five-step process for classroom observations that the observer must follow. This was developed to improve grading and inter-rater

reliability (Marzano & Toth, 2017). Each step requires the evaluator to select specific items or tasks to look for and then offers additional check boxes to identify techniques that have been implemented by the teacher. The steps are strategic, specific, and targeted to provide a fair and equitable evaluation of teacher behavior. The check boxes offer evaluators the opportunity to select predetermined indicators of effectiveness instead of having broad categories, as seen in older evaluation methods.

In step one, the evaluator observes the classroom teacher to identify scorable elements by asking the question "What elements am I seeing, and do they use the strategy correctly?" (Marzano & Toth, 2017) Evaluators can scroll through the menu to check any elements that are being implemented and then move to the next level and indicate whether the strategies are being utilized in that specific level (Marzano & Toth, 2017). Indicators must be present at the lowest levels before moving up to the next level.

Step two looks at how the teacher monitors the outcome. Once an evaluator identifies an element from instruction or conditions, they will observe the teacher and check a box for any monitoring techniques that are utilized during the observation (Marzano & Toth, 2017). The evaluator observes the classroom to determine whether the conditions for learning produce the desired effect (Marzano & Toth, 2017).

Next, the focus shifts from teacher to student and student work (Marzano & Toth, 2017). Step three asks for the specific percentage of students who are at the appropriate level of the target, and the observer must examine student work to

identify on-task behavior. The evaluator must determine the exact number of students who are achieving the desired outcome at the correct level. They do this by determining what percentage of students are working at the target level. If the evaluator observes less than 50% of students on task, the evaluator rates the teacher as "developing" (level 2). If 51%-90% of students are on task producing the desired effect, the evaluator moves the score to "applying" (level 3) (Marzano & Toth, 2017). If over 90% of the students are producing the desired effect, the teacher will move to "innovating" (level 4) (Marzano & Toth, 2017). The evaluator must understand the desired outcome to appropriately complete step three.

Step four requires the evaluator to identify ways in which the teacher modifies the lesson or behavior throughout the class period to produce the desired results. If students are not at the level of 90% producing the desired outcome, the evaluator must look for evidence that the teacher has adapted the lesson to help more students demonstrate the desired outcome (Marzano & Toth, 2017).

Step five uses student evidence to assign the final score for elements; this may include asking students questions or having the teacher bring evidence to the post-conference (Marzano & Toth, 2017). The evaluator may stop a student to ask where documents are kept, if they understand the daily objective, and what they believe the purpose of the lesson to be. This may also take place in a post-conference if necessary.

As previously mentioned, all steps are aligned to student performance and outcomes, creating a common vocabulary for principal and teacher conferences.

The steps provide encouragement for teachers who are selecting appropriate tasks for the classroom and provide strategic areas of improvement to aid in personal growth.

Aspects of the model

Research indicates that teachers must be able to provide feedback and input throughout the teacher evaluation process so that they are treated as learners instead of passive participants (Darling-Hammond et al., 2012). The MFTEM developers have taken this into consideration by adjusting the tool based on research and feedback since its initial launch in 2010. Initially, 41 elements were scored in the observation model, but that has been scaled back to 23 in the most current program (Marzano & Toth, 2017). The original domains included classroom strategies and behaviors, planning, preparing, reflecting on teaching, and collegiality. These domains had multiple sub-areas that were scored to measure teacher performance. However, these domains have been modified to meet the needs of current policy and streamline the evaluation process.

The domains under the MFTEM include standards-based planning, standards-based instruction, conditions for learning, and professional responsibilities. Within these four areas are 23 elements on which teachers receive a score at least once a year. These sub-areas include following: district policies and procedures, helping students revise knowledge, engaging in complex tasks, and examining reasoning. Each element requires teacher and student evidence, when applicable, to be used as part of the evaluation model. If evidence cannot be

found within the evaluation, teachers can clarify or provide additional evidence during the post-conference.

A level zero means that the element is not being used. Level one indicates "beginning," two "developing," three "applying," and four "innovating." Each score has a specific desired result listed below each level so that all evaluators utilize the same specific criteria to evaluate a teacher's performance. For example, on the scale for planning standards-based lessons, "not using" (level zero) means that the teacher "makes no attempt to plan rigorous units with learning targets embedded within a performance scale that demonstrates a progression of learning" (Marzano, 2017). "Beginning" (level one) asks whether the teacher is "using established content standards, attempts to plan rigorous units with learning targets embedded within a performance scale that demonstrates a progression of learning" (Marzano, 2017). "Developing" (level two) is similar to beginning except that the teacher is no longer merely *attempting* to plan but *actually planning* rigorous units. "Applying" (level 3) asks teachers to "provide evidence of implementing lessons/units & plans according to the level" (Marzano, 2017). To become an "innovative" teacher (level four), the teacher must "help others by sharing evidence of implementing lesson plans aligned to grade level standards using learning targets embedded in a performance scale and the impacts on student learning" (Marzano, 2017). The specific categories and rubrics allow for more rigorous and consistent scoring on teacher evaluations, eliminating the evaluator's potential bias in scoring. The MFTEM is then transferred to a numerical score so that it is equivalent to Oklahoma's five-tier rating system. The

point conversion is as follows: “not using” receives one point, “beginning” two points, “developing” three points, “applying” four points, and “innovating” five points. A teacher receives an evaluation multiple times a year. The lowest score is dropped, and the scores are averaged to provide a final score at the end of the school year. The areas of evaluation on the MFTEM line up with qualities of effective teaching identified by Strong, Ward, Tucker, and Hindman (2007). Key findings by Strong et al. (2007) include instruction, student assessment, classroom management, and personal qualities as indicators of effective, quality teachers. When teachers score higher in the four domains, their students demonstrate increased learning gains and fewer off-task behaviors.

In the selected urban school district, Teacher Leader Effectiveness District Guidelines for the 2018-2019 school year require that all qualifying teachers participate in two informal or unannounced observations and one formal announced observation. All of the observations, both informal and formal, count toward the final evaluation score. The informal observation should be 15-20 minutes long, and the teacher should not be informed that the observation will occur. The formal observation must be a minimum of 30 minutes long and include both a pre-conference and post-conference. The dates for the formal observation are set by the teacher and evaluator. Actionable feedback must be provided through iObservation within five days so that teachers can use this feedback to grow and increase effectiveness.

Prior to a scheduled formal observation, the teacher completes a pre-observation form so that the evaluator is familiar with the lesson objective, knows

what they should expect to see in the classroom, and understands the demographics of the classroom. The teachers are asked to identify the number of students who are English language learners and the number who receive additional service through an IEP. This puts less work on the evaluator, decreasing the time that was required under other observation tools, and helps explain class content prior to the evaluator's arrival. The pre-observation form also allows the teacher to go over the expectations of the teacher evaluation tool so that they are not surprised by what will be evaluated and how it will be scored. This facilitates growth-centered and actionable post-observation conferences that are specific to the teacher's classroom lesson.

When the observer arrives, they complete the evaluation using iObservation, noting evidence collected both visually and through student conversations. This information is then saved electronically and submitted to the teacher for review. A post-conference is scheduled so that the teacher can receive specific actionable feedback and provide additional evidence or support of student understanding that may be required. This type of feedback ensures that the teacher and observer are both engaged in the process in a fair and equitable way. MFTEM provides an opportunity for dialogue to happen between teachers and principals, which could increase the gains in student achievement. When principals and teachers work together as a team to create specific, actionable strategies, students' achievement scores are most likely to increase (Taylor & Tyler, 2012; Goe, Biggers, & Croft, 2012). At the conclusion of the post-conference, the teacher

should understand where they have specifically excelled in the classroom and should have developed a strategic, attainable goal for improvement.

Independent and Control Variables

The independent variables in this study include the teachers' certification level. Teacher certification includes standard, alternate, and emergency, as identified by the school district's administrative record. Teachers with standard certification have attended a traditional preparation program. Emergency-certified teachers have not attended a teacher preparation program prior to receiving the certificate. Alternatively certified teachers are employed as teachers while enrolled in a teacher preparation program.

The control variables included in this study are a teacher's gender, ethnicity, degree level, and the school level they were assigned to at the time of the evaluation. The school levels include elementary, middle school, and high school.

Data Analysis

The data was acquired by the school district as one complete data set that included teacher demographic data, licensing, highest degree, overall final evaluation score, and all 23 indicator scores. Initially the data set included 1484 participants: all staff members who received an evaluation for the 2018-19 school year. Librarians, counselors, and support staff were removed from the data set because they were not classroom teachers. This reduced the sample to 1285 teachers with completed evaluation scores and identified teaching certificates.

A linear regression was run to determine the association of the independent variable, teacher certification, and the dependent variable, final evaluation score. To control for differences in sociodemographic data, a series of regressions was completed. The second regression included controls for the binary variables gender, ethnicity, and degree level. The next model included the school level the teachers were assigned to during the school year.

An ordinary least square (OLS) regression was completed on the three indicators most frequently evaluated during the school year and included all possible scores. The subcategory indicators include (1) critical content, (2) rules and procedures and (3) policies and procedures. Initially, the OLS regression included teachers' evaluation score for the indicator area policies and procedures as the dependent variable and teacher certification level as the independent variable. To control for additional socioeconomic variables, an additional model included gender, ethnicity, and degree level. The third model included additional controls for the school level.

One might argue that the scale used for teacher evaluation falls under an ordinal classification rather than a nominal classification. To add robustness to this study, ordinal regression models were also investigated. Model 1 included the teachers' final evaluation score as the dependent variable and teacher certification as the independent variable. In the second model, controls were added to account for any association that might exist with respect to gender, ethnicity, and degree level. The third model included additional controls for the school level.

The next series of ordinal linear regression investigated the association between teachers' certification type and total evaluation score of teachers with scores in all three indicator areas. Model 1 included the total evaluation score as the dependent variable and teacher certification as the independent variable. In model 2, controls were added for gender, ethnicity, and degree level. The third model included additional controls for the school level. Additional ordinal linear regressions included the three indicators, individually running the same models for each.

I completed additional ordinal linear regression analysis limiting the sample to include only those with new teacher status. Boyd, Grossman, Lankford, Loeb, and Wyckoff (2009) found that first-year teachers' experiences and effectiveness were significantly different from those of second-year teachers. By removing career teachers, I was able to control for the additional years of classroom experience that may impact the teacher's evaluation score.

Initial ordinal regression models include the teachers' total evaluation score as the dependent variable and teacher certification level as the independent variable. In model 2, controls were added for gender, ethnicity, and degree level. The third model included additional controls for the school level. Ordinal regressions models also included the three evaluation scores for the indicator areas (1) critical content, (2) rules and procedures and (3) policies and procedures. Model 1 included only the certification type and evaluation score. Model 2 added additional variables including gender, ethnicity, and degree level. The third model included the addition of school level.

Chapter 4: Results

The Data

The original data set included n =1484 evaluation scores. Participants were excluded if they were not classroom teachers or they did not have a teaching certificate. The remaining 1285 participants were included in the analysis because they were classroom teachers with a teaching certificate and evaluation scores. To assess whether the type of teaching certificate predicts a teacher's evaluation score, I ran a linear regression on the overall evaluation score. The 23 individual domains were then analyzed to determine which domains contained the largest number of entries: (1) critical content, (2) rules and procedures and (3) policies and procedures .

This study explored the association between total evaluation scores and teachers' certification level and examined how several independent variables, including sociodemographic data, might affect the total evaluation score. I was guided by the hypotheses below:

H^1: A difference exists in teacher quality between teachers who hold traditional teacher certificates and those who hold alternative certificates.

H^2: A difference in teacher quality exists between new teachers who hold a traditional teacher certificate and those who hold an emergency or alternative teaching certificate.

Data analysis will be run in SPSS to test the differences in quality of

This chapter presents the data analysis including ordinary least squares (OLS) regression and robustness checks by hypothesis.

Table 2. The bivariate correlation of dependent variables and teachers' certification type

	Total Evaluation Score	Critical Content	Rules & Procedures	Policies & Procedures	Standard	Emergency	Alternate
Total Evaluation Score	--						
	1285						
Critical Content	.696**	--					
	0.000						
	1209	1209					
Rules & Procedures	.632**	.397**	--				
	0.000	0.000					
	1045	988	1045				
Policies & Procedures	.599**	.366**	.310**	--			
	0.000	0.000	0.000				
	1156	1094	971	1156			
Standard	.136**	0.055	.140**	0.027	--		
	0.000	0.055	0.000	0.356			
	1285	1209	1045	1156	1285		
Emergency	-.187**	-.109**	-.171**	-.058*	-.691**	--	
	0.000	0.000	0.000	0.047	0.000		
	1285	1209	1045	1156	1285	1285	
Alternate	0.033	0.051	0.010	0.031	-.545**	-.230**	--
	0.234	0.077	0.747	0.285	0.000	0.000	
	1285	1209	1045	1156	1285	1285	1285

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Results

H^1: A difference exists in teacher quality between teachers who hold traditional teacher certificates and those who hold alternative certificates.

To test H^1, a linear regression analysis was conducted to evaluate the relationship of teacher evaluation score and teacher certification level.

Table 3. OLS Regression examining the relationship between teachers' certification status and total evaluation score

	(1)	(2)	(3)
VARIABLES	**Total Evaluation Score**	**Total Evaluation Score**	**Total Evaluation Score**
Emergency	-0.22***	-0.20***	-0.19***
	(0.03)	(0.04)	(0.04)
Alternate	-0.01	-0.00	-0.01
	(0.04)	(0.04)	(0.04)
Gender		0.15***	0.16***
		(0.03)	(0.03)
Black		-0.06	-0.07
		(0.04)	(0.04)
Hispanic Latino/a		-0.02	-0.01
		(0.05)	(0.05)
Other		-0.04	-0.04
		(0.05)	(0.05)
Degree Level		0.02	0.03
		(0.03)	(0.03)
High School			0.07*
			(0.03)
Middle School			0.00
			(0.04)
Constant	3.63***	3.52***	3.48***
	(0.02)	(0.03)	(0.04)
Observations	1,285	1,285	1,285
Adjusted R-squared	0.03	0.05	0.05

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

In model 1, I introduced binary variables indicating certification status. Relative to standard certification status, emergency certification status is negatively associated with total evaluation score ($p < .001$). The magnitude of the negative association is -0.22. Alternative certification status is not statistically associated with total evaluation score when compared to standard certification status. In Model 2, teachers' sociodemographic background characteristics are introduced for gender, race, and degree level. With the introduction of these variables, the negative association between certification status remains about the same, at -0.20, and is statistically significant ($p < .001$). Additionally, alternative certification status is not statistically associated with total evaluation score after teachers' sociodemographic background factors are entered into the model. With regard to gender, being female is associated with higher total evaluation scores than being male. The positive association is 0.15 and is statistically significant ($p < .001$). The remaining background characteristic are not statistically associated with total evaluation scores. In Model 3, controls for school level are added, with similar results. Relative to standard certification status, emergency certification is still negatively associated with total evaluation score ($p < .001$). The magnitude of the negative association is -0.19. Compared to standard certification status, alternative certification status is not statistically associated with total evaluation score. The positive association is 0.16 and is statistically significant ($p < .001$). The remaining background characteristics are not statistically associated with total evaluation scores. For school level, high school is positively associated

with total evaluation score relative to elementary school, but middle school is not associated with total evaluation scores compared to elementary school.

Table 4. OLS regression models investigating the relationship between teacher certification and the critical content sub-score indicator

VARIABLES	(1) Critical Content	(2) Critical Content	(3) Critical Content
Emergency	-0.18***	-0.16**	-0.15**
	(0.05)	(0.05)	(0.05)
Alternate	0.06	0.06	0.05
	(0.06)	(0.06)	(0.06)
Gender		0.17***	0.20***
		(0.05)	(0.05)
Black		-0.04	-0.04
		(0.06)	(0.06)
Hispanic Latino		-0.01	-0.01
		(0.08)	(0.08)
Other		-0.01	-0.01
		(0.08)	(0.08)
Degree level		0.04	0.05
		(0.04)	(0.05)
Highschool			0.07
			(0.05)
Middle school			0.09
			(0.06)
Constant	3.60***	3.45***	3.39***
	(0.03)	(0.05)	(0.06)
Observations	1,209	1,209	1,209
Adjusted R-squared	0.01	0.02	0.02

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

In model 1, I introduced binary variables indicating certification status. Relative to standard certification status, emergency certification status is negatively associated with critical content score ($p < .001$). The magnitude of the negative association is -0.18. Alternative certification status is not statistically

associated with critical content score when compared to standard certification status. In Model 2, teachers' sociodemographic background characteristics are introduced for gender, race, and degree level. With the introduction of these variables, the negative association between certification status remains about the same at -0.16 and is statistically significant ($p < .001$). Alternative certification status is not statistically associated with critical content score after teachers' sociodemographic background factors are entered into the model. With regard to gender, being female is associated with higher critical content scores than being male. The positive association is 0.17 and is statistically significant ($p < .001$). The additional background characteristics are not statistically associated with critical content scores. In Model 3, controls for school level are added. Results remain largely similar. Relative to standard certification status, emergency certification is still negatively associated with critical content score ($p < .001$). The magnitude of the negative association is -0.15. Alternative certification status is again not statistically associated with critical content score when compared to standard certification status. The positive association is 0.16 and is statistically significant ($p < .001$). The remaining background characteristic are not statistically associated with critical content scores. For school level, high school is positively associated with critical content score relative to elementary school, but middle school is not associated with total evaluation scores compared to elementary school. However, this positive association is not statistically significant.

Table 5. OLS regression models investigating the relationship between teacher certification and the sub-score indicator, establishing, and acknowledging adherence to rules and procedures

VARIABLES	(1) Rules & Procedures	(2) Rules & Procedures	(3) Rules & Procedures
Emergency	-0.29***	-0.29***	-0.30***
	(0.05)	(0.05)	(0.06)
Alternate	-0.06	-0.04	-0.03
	(0.06)	(0.06)	(0.06)
Gender		0.13**	0.10*
		(0.05)	(0.05)
Black		0.04	0.04
		(0.06)	(0.06)
Hispanic Latino/a		-0.04	-0.03
		(0.08)	(0.08)
Other		-0.06	-0.05
		(0.08)	(0.08)
Degree Level		-0.04	-0.05
		(0.05)	(0.05)
High School			-0.01
			(0.05)
Middle School			-0.18**
			(0.06)
Constant	3.83***	3.77***	3.82***
	(0.03)	(0.05)	(0.06)
Observations	1,045	1,045	1,045
Adjusted R-squared	0.03	0.03	0.04

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

In model 1, I introduced binary variables indicating certification status. Relative to standard certification status, emergency certification status is negatively associated with rules and procedures score ($p < .001$). The magnitude of the negative association is -0.29. Alternative certification status is not statistically associated with rules and procedures scores when compared to standard certification status. In Model 2, teachers' sociodemographic background

characteristics are introduced for gender, race, and degree level. With the introduction of these variables, the negative association between certification status remains about the same at -0.29 and is statistically significant ($p < .001$). Alternative certification status is not statistically associated with rules and procedures score after teachers' sociodemographic background factors are entered into the model. With regard to gender, being female is associated with higher rules and procedures scores than being male. The positive association is 0.13 and is statistically significant ($p < .001$). The additional background characteristics are not statistically associated with rules and procedures scores. In Model 3, controls for school level are added. Results remain largely similar. Relative to standard certification status, emergency certification is still negatively associated with rules and procedures scores ($p < .001$). The magnitude of the negative association is -0.30. Alternative certification status is not statistically associated with rules and procedures score when compared to standard certification status. The positive association of gender is 0.10 and is statistically significant ($p < .05$). The sociodemographic information continues to have no statistically significant association with teachers' rules and procedures scores. For school level, middle school is positively associated with rules and procedures scores relative to elementary school. This positive association is statistically significant ($p < .01$).

Table 6. OLS regression investigating the association of teachers' total evaluation score and sub-score indicator, policies and procedures

VARIABLES	(1) Policies & Procedures	(2) Policies & Procedures	(3) Policies & Procedures
Emergency	-0.09	-0.06	-0.06
	(0.05)	(0.05)	(0.05)
Alternative	0.04	0.04	0.05
	(0.06)	(0.06)	(0.06)
Gender		0.12**	0.11*
		(0.05)	(0.05)
Black		-0.09	-0.09
		(0.05)	(0.05)
Hispanic Latino/a		0.05	0.05
		(0.08)	(0.08)
Other		0.01	0.01
		(0.08)	(0.08)
Degree Level		0.02	0.02
		(0.04)	(0.04)
High school			-0.03
			(0.05)
Middle school			-0.04
			(0.06)
Constant	3.71***	3.62***	3.64***
	(0.03)	(0.05)	(0.06)
Observations	1,156	1,156	1,156
Adjusted R-squared	0.00	0.01	0.01

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

In model 1, I introduced binary variables indicating certification status. Relative to standard certification status, emergency certification status is negatively associated with policies and procedures scores but is not statistically significant. The magnitude of the negative association is -0.09. Alternative certification status continues to have no statistically significant association with total evaluation score when compared to standard certification status. In Model 2,

teachers' sociodemographic background characteristics are introduced for gender, race, and degree level. With the introduction of these variables, the negative association between certification status remains about the same at -0.06 and is not statistically significant. Alternative certification status is not statistically associated with policies and procedures scores after teachers' sociodemographic background factors are entered into the model. With regard to gender, being female is associated with higher policies and procedures scores than being male the positive association is 0.12 and is statistically significant ($p < .01$). The additional background characteristics are not statistically associated with policies and procedures scores. In Model 3, controls for school level are added. Results remain similar across models. Relative to standard certification status, emergency certification is still negatively associated with policies and procedures scores. The magnitude of the negative association is -0.06. Alternative certification status is not statistically associated with policies and procedures scores when compared to standard certification status. The positive association of gender is 0.11 and continues to be statistically significant ($p < .01$). The sociodemographic information continues to have no statistically significant association with teachers' policies and procedures scores. For school level, no significant associations exist with total evaluations score when compared to elementary schools.

In the next series of models, robustness checks are performed. In the table 7, the results of an ordinal regression are presented to test the robustness of the main OLS regression models.

Table 7. Robust ordinal regression investigating the association of teachers' certification and total evaluation score

VARIABLES	(1) Total Evaluation Score	(2) Total Evaluation Score	(3) Total Evaluation Score
Emergency	-0.80***	-0.68***	-0.65***
	(0.12)	(0.13)	(0.13)
Alternate	-0.07	-0.02	-0.03
	(0.14)	(0.14)	(0.14)
Gender		0.51***	0.56***
		(0.11)	(0.12)
Black		-0.31*	-0.32*
		(0.13)	(0.13)
Hispanic Latino/a		-0.09	-0.08
		(0.19)	(0.19)
Other		-0.22	-0.22
		(0.17)	(0.17)
Degree Level		0.03	0.05
		(0.11)	(0.11)
High School			0.19
			(0.12)
Middle School			-0.01
			(0.15)
Observations	1,285	1,285	1,285
r2_p	0.00604	0.00975	0.0101

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

In model 1, I introduced binary variables indicating certification status. Relative to standard certification status, emergency certification status is negatively associated with total evaluation score ($p < .001$). The magnitude of the negative association is -0.80. Alternative certification status is not statistically associated with total evaluation score when compared to standard certification status. In Model 2, teachers' sociodemographic background characteristics are introduced for gender, race, and degree level. With the introduction of these

variables, the negative association between certification status reduces to -0.68 and is statistically significant ($p < .001$). Alternative certification status is not statistically associated with total evaluation score after teachers' sociodemographic background factors are entered into the model. With regard to gender, being female is associated with higher total evaluation scores than being male. The positive association is 0.51 and is statistically significant ($p < .001$). With regard to race, being Black is negatively associated with total evaluation scores when compared to being white, with a magnitude of -0.31 ($p < .05$). The additional background characteristics are not statistically associated with total evaluation scores. In Model 3, controls for school level are added. Results remain similar to Model 2. Relative to standard certification status, emergency certification is still negatively associated with total evaluation score ($p < .001$). The magnitude of the negative association is -0.65. Alternative certification status is again not statistically associated with total evaluation score when compared to standard certification status. The positive association of gender is 0.56 and is statistically significant ($p < .001$). The sociodemographic variables continue to have little statistically significant association with teachers' total evaluation scores, except for the race variable; being Black is negatively associated with total evaluation scores when compared to being white, with a magnitude of -0.32 ($p < .05$). For school level, middle school is negatively associated with total evaluation score relative to elementary school but is not statistically significant.

Table 8. Robust ordinal regression results exploring the association of teachers' certification and sub-score indicator, critical content

VARIABLES	(1) Critical Content	(2) Critical Content	(3) Critical Content
Emergency	-0.50***	-0.46**	-0.43**
	(0.14)	(0.15)	(0.15)
Alternate	0.10	0.14	0.09
	(0.16)	(0.16)	(0.17)
Gender		0.46***	0.54***
		(0.13)	(0.14)
Black		-0.13	-0.13
		(0.15)	(0.15)
Hispanic Latino/a		-0.01	0.01
		(0.22)	(0.22)
Other		-0.03	-0.03
		(0.21)	(0.21)
Degree Level		0.08	0.11
		(0.12)	(0.12)
High School			0.20
			(0.14)
Middle School			0.28
			(0.18)
Observations	1,209	1,209	1,209
r2_p	0.00628	0.0121	0.0135

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

In model 1, I introduced binary variables indicating certification status of teachers. Relative to standard certification status, emergency certification status is negatively associated with critical content score ($p < .001$), with a magnitude of -0.50. Alternative certification status is not statistically associated with critical content score when compared to standard certification status. In Model 2, teachers' sociodemographic background characteristics are introduced for gender, race, and degree level. With the introduction of these variables, the negative

association between certification status remains about the same at -0.46 and is statistically significant ($p < .01$). Alternative certification status is not statistically associated with critical content score after teachers' sociodemographic background factors are entered into the model. With regard to gender, being female is associated with higher critical content scores than being male. The positive association is 0.46 and is statistically significant ($p < .001$). The additional background characteristics are not statistically associated with critical content scores. In Model 3, controls for school level are added. Results remain similar. Relative to standard certification status, emergency certification is still negatively associated with critical content score with a magnitude of -0.43 ($p < .01$) when compared to standard certification status; alternative certification status is not statistically associated with critical content score. The positive association of gender is 0.54 and is statistically significant ($p < .001$). The sociodemographic information continues to have no statistically significant association with teachers' critical content scores. For school level, no statistically significant association exists in Model 3 when comparing elementary critical content scores to middle school or high school critical content scores.

Table 9. Robust ordinal regression results to investigate the association of teachers' certificate and sub-score indicator, rules and procedures

VARIABLES	(1) Rules & Procedures	(2) Rules & Procedures	(3) Rules & Procedures
Emergency	-0.88***	-0.88***	-0.91***
	(0.15)	(0.16)	(0.17)
Alternate	-0.18	-0.13	-0.09
	(0.18)	(0.19)	(0.19)
Gender		0.33*	0.24
		(0.14)	(0.15)
Black		0.11	0.10
		(0.17)	(0.17)
Hispanic Latino/a		-0.09	-0.08
		(0.25)	(0.25)
Other		-0.18	-0.16
		(0.22)	(0.23)
Degree Level		-0.14	-0.17
		(0.14)	(0.14)
High school			-0.10
			(0.16)
Middle school			-0.43*
			(0.20)
Observations	1,045	1,045	1,045
r2_p	0.0162	0.0198	0.0221

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

In model 1, I introduced binary variables indicating certification status. Relative to standard certification status, emergency certification status is negatively associated with rules and procedures scores ($p < .001$). The magnitude of the negative association is -0.88. Alternative certification status is not statistically associated with rules and procedures scores when compared to standard certification status. In Model 2, teachers' sociodemographic background characteristics are introduced for gender, race, and degree level. With the

introduction of these variables, the negative association between certification status remains the same at -0.88 and is statistically significant ($p < .001$). Alternative certification status is not statistically associated with rules and procedures score after teachers' sociodemographic background factors are entered into the model. With regard to gender, being female is associated with higher rules and procedures scores than being male. The positive association is 0.33 and is statistically significant ($p < .05$). The additional background characteristics are not statistically associated with rules and procedures scores. In Model 3, controls for school level are added. Results remain similar across the models. Relative to standard certification status, emergency certification is still negatively associated with rules and procedures score ($p < .001$). The magnitude of the negative association is -0.91. Alternative certification status is again not statistically associated with rules and procedures score when compared to standard certification status. The positive association of gender is 0.24 and is not statistically significant. The sociodemographic information continues to have no statistically significant association with teachers' rules and procedures scores. For school level, middle school is negatively associated with rules and procedures score relative to elementary school. This negative association is statistically significant ($p < .05$) with a magnitude of -0.43.

Table 10. Robust ordinal regression results that explore the association of teachers' certification and the sub-score indicator, policies and procedures

VARIABLES	(1) Policies & Procedures	(2) Policies & Procedures	(3) Policies & Procedures
Emergency	-0.22	-0.12	-0.14
	(0.14)	(0.15)	(0.15)
Alternate	0.14	0.16	0.17
	(0.17)	(0.17)	(0.17)
Gender		0.34*	0.31*
		(0.13)	(0.14)
Black		-0.29	-0.28
		(0.16)	(0.16)
Hispanic Latino/a		0.07	0.06
		(0.24)	(0.24)
Other		-0.05	-0.04
		(0.21)	(0.21)
Degree Level		0.03	0.02
		(0.13)	(0.13)
High School			-0.11
			(0.15)
Middle School			-0.04
			(0.18)
Observations	1,156	1,156	1,156
r2_p	0.00159	0.00630	0.00652

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

In model 1, I introduced binary variables indicating teacher certification status. Relative to standard teacher certification status, emergency certification status is negatively associated with policies and procedures score but is not statistically significant. The magnitude of the negative association -0.22. Alternative certification status is not statistically associated with policies and procedures score when compared to standard teacher certification status. In Model 2, teachers' sociodemographic background characteristics are introduced for

gender, race, and degree level. With the introduction of these variables, the negative association between certification changes to -0.12 and is still not statistically significant. Alternative certification status is not statistically associated with policies and procedures score after teachers' sociodemographic background factors are entered into the model. With regard to gender, being female is associated with higher policies and procedures scores than being male. The positive association is 0.34 and is statistically significant ($p < .05$). The additional background characteristics introduced into the model are not statistically associated with policies and procedures scores. In Model 3, controls for school level are added. Results remain similar to Model 2. Relative to standard certification status, emergency certification is still negatively associated with policies and procedures score. The magnitude of the negative association is -0.14. Alternative certification status is not statistically associated with policies and procedures score when compared to standard certification status. The positive association of gender is 0.31 and is statistically significant ($p < .05$). The sociodemographic information continues to have no statistically significant association with teachers' policies and procedures scores. For school level, middle school has no significant association with policies and procedures score when compared to elementary schools.

In the next series of models, regressions are completed and robustness checks are performed, but the sample is restricted to teachers who have a score for each outcome variable indicated as a higher priority by the school district.

Table 11. OLS regression investigating the association between teachers' certification type and total evaluation score teachers who had scores in all three indicators

VARIABLES	(1) Total Evaluation Score	(2) Total Evaluation Score	(3) Total Evaluation Score
Emergency	-0.23***	-0.20***	-0.20***
	(0.04)	(0.04)	(0.04)
Alternate	-0.02	-0.01	-0.01
	(0.04)	(0.04)	(0.05)
Gender		0.16***	0.16***
		(0.03)	(0.04)
Black		-0.06	-0.06
		(0.04)	(0.04)
Hispanic Latino/a		0.01	0.02
		(0.06)	(0.06)
Other		-0.03	-0.03
		(0.06)	(0.06)
Degree Level		0.01	0.01
		(0.03)	(0.03)
High School			0.06
			(0.04)
Middle School			-0.07
			(0.05)
Constant	3.63***	3.52***	3.51***
	(0.02)	(0.04)	(0.05)
Observations	925	925	925
Adjusted R-squared	0.04	0.06	0.06

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

In model 1, I introduced binary variables indicating certification status. Relative to standard certification status, emergency certification status is negatively associated with total evaluation score ($p < .001$). The magnitude of the negative association is -0.23. Alternative certification status is not statistically associated with total evaluation score when compared to standard certification status. In Model 2, teachers' sociodemographic background characteristics are introduced for gender, race, and degree level. With the introduction of these

variables, the negative association between certification status remains about the same at -0.20 and is statistically significant ($p < .001$). Alternative certification status is not statistically associated with total evaluation score after teachers' sociodemographic background factors are entered into the model. With regard to gender, being female is associated with higher total evaluation scores than being male. The positive association is 0.16 and is statistically significant ($p < .001$). The additional background characteristics are not statistically associated with total evaluation scores. In Model 3, controls for school level are added. Results remain largely similar across models. Relative to standard certification status, emergency certification is still negatively associated with total evaluation score ($p < .001$). The magnitude of the negative association is -0.20. Alternative certification status is again not statistically associated with total evaluation score when compared to standard certification status. The positive association of gender is 0.16 and continues to be statistically significant ($p < .001$). The sociodemographic information continues to have no statistically significant association with teachers' total evaluation scores. For school level, middle school is negatively associated with total evaluation score relative to elementary school and the relationship is not statistically significant.

Table 12. OLS regression investigating the association of teachers' certification status and subcategory indicator, critical content, of teachers who had scores in all three indicators

VARIABLES	(1) Critical Content	(2) Critical Content	(3) Critical Content
Emergency	-0.19***	-0.16*	-0.15*
	(0.06)	(0.06)	(0.06)
Alternate	0.06	0.07	0.06
	(0.07)	(0.07)	(0.07)
Gender		0.19***	0.20***
		(0.05)	(0.06)
Black		-0.08	-0.08
		(0.06)	(0.06)
Hispanic Latino/a		0.04	0.04
		(0.09)	(0.09)
Other		-0.01	-0.02
		(0.09)	(0.09)
Degree Level		0.01	0.02
		(0.05)	(0.05)
High School			0.04
			(0.06)
Middle School			0.01
			(0.07)
Constant	3.60***	3.46***	3.43***
	(0.03)	(0.06)	(0.07)
Observations	925	925	925
Adjusted R-squared	0.01	0.02	0.02

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

In model 1, I introduced binary variables indicating certification status. Relative to standard certification status, emergency certification status is negatively associated with critical content score ($p < .001$). The magnitude of the negative association is -0.19. Alternative certification status is not statistically associated with total evaluation score when compared to standard certification status. In Model 2, teachers' sociodemographic background characteristics are

introduced for gender, race, and degree level. With the introduction of these variables, the negative association between certification status remains about the same at -0.16 and is statistically significant ($p < .001$). Alternative certification status is not statistically associated with critical content score after teachers' sociodemographic background factors are entered into the model. With regard to gender, being female is associated with higher critical content scores than being male. The positive association is 0.19 and is statistically significant ($p < .001$). The additional background characteristics are not statistically associated with total evaluation scores. In Model 3, controls for school level are added. Results remain largely similar across models. Relative to standard certification status, emergency certification is still negatively associated with total evaluation score ($p < .001$). The magnitude of the negative association is -0.15. Alternative certification status is again not statistically associated with critical content score when compared to standard certification status. The positive association of gender is 0.20 and continues to be statistically significant ($p < .01$). The sociodemographic information continues to have no statistically significant association with teachers' critical content scores. For school level, middle school is no longer negatively associated with total evaluation score relative to elementary school. The relationship is not statistically significant.

Table 13. OLS regression results investigating the association of teachers' certification and subcategory indicator, rules and procedures when the teacher had scores in all three indicators

VARIABLES	(1) Rules & Procedures	(2) Rules & Procedures	(3) Rules & Procedures
Emergency	-0.29***	-0.30***	-0.31***
	(0.05)	(0.06)	(0.06)
Alternate	-0.07	-0.06	-0.04
	(0.06)	(0.06)	(0.07)
Gender		0.12*	0.09
		(0.05)	(0.05)
Black		0.08	0.08
		(0.06)	(0.06)
Hispanic Latino/a		-0.02	-0.01
		(0.09)	(0.09)
Other		-0.02	-0.01
		(0.08)	(0.08)
Degree Level		-0.03	-0.04
		(0.05)	(0.05)
High School			-0.02
			(0.06)
Middle School			-0.17*
			(0.07)
Constant	3.84***	3.75***	3.81***
	(0.03)	(0.06)	(0.07)
Observations	925	925	925
Adjusted R-squared	0.03	0.03	0.04

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

In model 1, I introduced binary variables indicating certification status. Relative to standard certification status, emergency certification status is negatively associated with rules and procedures score ($p < .001$). The magnitude of the negative association is -0.29. Alternative certification status is not statistically associated with rules and procedures score when compared to

standard certification status. In Model 2, teachers' sociodemographic background characteristics are introduced for gender, race, and degree level. With the introduction of these variables, the negative association between certification status remains about the same at -0.30 and is statistically significant ($p < .001$). Alternative certification status is not statistically associated with rules and procedures score after teachers' sociodemographic background factors are entered into the model. With regard to gender, being female is associated with higher rules and procedures scores than being male. The positive association is 0.12 and is statistically significant ($p < .05$). The additional background characteristics are not statistically associated with rules and procedures scores. In Model 3, controls for school level are added. Results remain largely similar across models. Relative to standard certification status, emergency certification is still negatively associated with rules and procedures score ($p < .001$). The magnitude of the negative association is -0.31. Alternative certification status is again not statistically associated with rules and procedures score when compared to standard certification status. The positive association of gender is 0.09 but is no longer statistically significant. The sociodemographic information continues to have no statistically significant association with teachers' rules and procedures scores. Middle school is negatively associated with rules and procedures score relative to elementary school at a magnitude of -0.17 and is statistically significant ($p < .05$).

Table 14. OLS regression results investigating the association of teachers' certification and subcategory indicator, policies and procedures for the teacher had scores in all three indicators

VARIABLES	(1) Policies & Procedures	(2) Policies & Procedures	(3) Policies & Procedures
Emergency	-0.09	-0.06	-0.06
	(0.06)	(0.06)	(0.06)
Alternate	0.04	0.04	0.05
	(0.07)	(0.07)	(0.07)
Gender		0.13*	0.12*
		(0.05)	(0.06)
Black		-0.08	-0.08
		(0.06)	(0.06)
Hispanic Latino/a		0.02	0.02
		(0.09)	(0.09)
Other		0.02	0.02
		(0.08)	(0.08)
Degree Level		0.01	0.01
		(0.05)	(0.05)
High School			-0.01
			(0.06)
Middle School			-0.08
			(0.07)
Constant	3.71***	3.61***	3.64***
	(0.03)	(0.06)	(0.07)
Observations	925	925	925
Adjusted R-squared	0.00	0.01	0.01

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

In model 1, I introduced binary variables indicating certification status. Relative to standard certification status, emergency certification status is negatively associated with policies and procedures score ($p < .001$). The magnitude of the negative association is -0.09. Alternative certification status is not statistically associated with policies and procedures score when compared to standard certification status. In Model 2, teachers' sociodemographic background

characteristics are introduced for gender, race, and degree level. With the introduction of these variables, the negative association between certification status magnitude is -0.06 and is not statistically significant. Alternative certification status is not statistically associated with policies and procedures score after teachers' sociodemographic background factors are entered into the model. With regard to gender, being female is associated with higher policies and procedures scores than being male. The positive association is 0.13 and is statistically significant ($p < .05$). The additional background characteristics are not statistically associated with policies and procedures scores. In Model 3, controls for school level are added. Results remain largely similar across models. Relative to standard certification status, emergency certification is still negatively associated with policies and procedures score but is not statistically significant. The magnitude of the negative association is -0.06. Alternative certification status is not statistically associated with policies and procedures score when compared to standard certification status. The positive association of gender is 0.12 ($p < .05$). The sociodemographic information continues to have no statistically significant association with teachers' policies and procedures scores. Relative to elementary status, high school and middle school are not significantly associated with policies and procedures scores.

H^2: A difference in teacher quality exists between new teachers who hold a traditional teacher certificate and those who hold an emergency or alternative teaching certificate

In the previous models, both new and career teachers' certification status were investigated. In the following models, only new teachers are considered for analysis. Career teachers are those who hold a standard teaching certificate and have worked in the school district for three or more consecutive complete school years.

Table 15. Results of OLS regression models investigating the relationship between teacher certification and total evaluation score for only new teachers.

	(1)	(2)	(3)
VARIABLES	**Total Evaluation Score**	**Total Evaluation Score**	**Total Evaluation Score**
Emergency	-0.28***	-0.25***	-0.24***
	(0.04)	(0.04)	(0.04)
Alternate	-0.10*	-0.09*	-0.09*
	(0.04)	(0.04)	(0.04)
Gender		0.16***	0.17***
		(0.04)	(0.04)
Black		-0.08*	-0.08*
		(0.04)	(0.04)
Hispanic Latino/a		0.00	0.01
		(0.06)	(0.06)
Other		-0.05	-0.05
		(0.05)	(0.05)
Degree Level		-0.00	0.00
		(0.04)	(0.04)
High School			0.06
			(0.04)
Middle School			-0.01
			(0.05)
Constant	3.68***	3.58***	3.55***
	(0.02)	(0.04)	(0.05)
Observations	882	882	882
Adjusted R-squared	0.06	0.08	0.08

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

In model 1, I introduced binary variables indicating certification status. Relative to standard certification status for new teachers, emergency certification status is negatively associated with total evaluation score ($p < .001$). The magnitude of the negative association is -0.28. Alternative certification status is statistically associated with total evaluation score when compared to standard certification status ($p < .05$). The negative association is -0.10. In Model 2, teachers' sociodemographic background characteristics are introduced for gender, race, and degree level. With the introduction of these variables, the negative association between certification status remains about the same at -0.25 and is statistically significant ($p < .001$). Additionally, alternative certification status is statistically associated with total evaluation score after teachers' sociodemographic background factors are entered into the model with a magnitude of -0.09 ($p < .05$). With regard to gender, being female is associated with higher total evaluations scores than being male, with a positive association of magnitude 0.16 ($p < .001$). With regard to race, being Black is associated with lower total evaluation scores than being white, with a magnitude of -0.08 ($p < .05$). In Model 3, controls for school level are added. Results remain largely similar. Relative to standard certification status, emergency certification is still negatively associated with total evaluation score ($p < .001$). The magnitude of the negative association is -0.24. Alternative certification status is statistically associated with total evaluation score when compared to standard certification status with a magnitude of -0.09 ($p < .05$). Relative to males, females have a positive significant association with total evaluation scores with a magnitude of

0.17 ($p < .001$). The background characteristic, Black, is significantly associated with total evaluation score when compared to white. The magnitude of the association is -0.08($p < .05$). The remaining background characteristic are not statistically associated with total evaluation scores. Relative to elementary status, high school and middle school are not significantly associated with total evaluation scores.

Table 16. OLS regression table investigating the relationship between teacher certification and sub-score indicator, critical content, for new teachers only

VARIABLES	(1) Critical Content	(2) Critical Content	(3) Critical Content
Emergency	-0.26***	-0.21***	-0.20***
	(0.06)	(0.06)	(0.06)
Alternate	-0.06	-0.05	-0.06
	(0.07)	(0.07)	(0.07)
Gender		0.18**	0.21***
		(0.05)	(0.06)
Black		-0.11	-0.11
		(0.06)	(0.06)
Hispanic Latino/a		0.05	0.05
		(0.10)	(0.10)
Other		-0.06	-0.06
		(0.08)	(0.08)
Degree Level		0.01	0.02
		(0.05)	(0.05)
High School			0.06
			(0.06)
Middle School			0.09
			(0.07)
Constant	3.67***	3.55***	3.49***
	(0.03)	(0.06)	(0.08)
Observations	833	833	833
Adjusted R-squared	0.02	0.04	0.04

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

In model 1, I introduced binary variables indicating new teacher certification status. Relative to standard certification status, emergency certification status is negatively associated with critical content score ($p < .001$). The magnitude of the negative association is -0.26. Alternative certification status is not statistically associated with critical content score when compared to standard certification status. In Model 2, teachers' sociodemographic background characteristics are introduced for gender, race, and degree level. With the introduction of these variables, the negative association between certification status magnitude is -0.21 ($p < .001$). Alternative certification status is not statistically associated with critical content score after teachers' sociodemographic background factors are entered into the model. With regard to gender, being female is associated with higher total critical content scores than being male. The positive association is 0.18 and is statistically significant ($p < .001$). The additional background characteristics are not statistically associated with critical content scores. In Model 3, controls for school level are added. Results remain largely similar across models. Relative to standard certification status, emergency certification is still negatively associated with critical content score and remains statistically significant ($p < .001$). The magnitude of the negative association is -0.20. Alternative certification status is not statistically associated with critical content score when compared to standard certification status. The positive association of gender is 0.21 ($p < .001$). The remaining sociodemographic information continues to have no statistically significant

association with teachers' critical content scores. Relative to elementary status, high school and middle school are not significantly associated with critical content scores.

Table 17. OLS regression table investigating the relationship between certification status and sub-score indicator, rules and procedures, for new teachers

VARIABLES	(1) Rules & Procedures	(2) Rules & Procedures	(3) Rules & Procedures
Emergency	-0.32***	-0.33***	-0.34***
	(0.06)	(0.06)	(0.06)
Alternate	-0.17*	-0.15*	-0.14
	(0.07)	(0.07)	(0.07)
Gender		0.20***	0.16**
		(0.06)	(0.06)
Black		0.06	0.06
		(0.07)	(0.07)
Hispanic Latino/a		-0.02	-0.01
		(0.10)	(0.10)
Other		0.00	0.02
		(0.09)	(0.09)
Degree Level		-0.03	-0.03
		(0.06)	(0.06)
High School			-0.02
			(0.07)
Middle School			-0.20*
			(0.08)
Constant	3.86***	3.72***	3.78***
	(0.04)	(0.07)	(0.08)
Observations	726	726	726
Adjusted R-squared	0.04	0.05	0.05

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

In model 1, I introduced binary variables indicating new teacher certification status. Relative to standard certification status, emergency certification status is negatively associated with rules and procedures score ($p <$

.001). The magnitude of the negative association is -0.32. Alternative certification status is statistically associated with rules and procedures score when compared to standard certification status ($p < .001$). In Model 2, teachers' sociodemographic background characteristics are introduced for gender, race, and degree level. With the introduction of these variables, the negative association between certification status magnitude is -0.33 and is statistically significant ($p < .001$). Alternative certification status is statistically associated with rules and procedures by a magnitude of -0.15 ($p < .05$). With regard to gender, being female is associated with higher rules and procedures scores than being male. The positive association is 0.20 and is statistically significant ($p < .001$). The additional background characteristics are not statistically associated with rules and procedures scores. In Model 3, controls for school level are added. Results remain largely similar across models. Relative to standard certification status, emergency certification is still negatively associated with rules and procedures with a magnitude of -0.14. Alternative certification status is not statistically associated with rules and procedures score when compared to standard certification status. The positive association of gender is 0.16 ($p < .01$). The sociodemographic information continues to have no statistically significant association with teachers' rules and procedures scores. Relative to elementary status, high school is not significantly associated with rules and procedures scores. However, middle school is significantly associated with a magnitude of -0.20 ($p < .05$).

Table 18. OLS regression investigating the relationship of new teacher certification and sub-score indicator, policies and procedures

VARIABLES	(1) Policies & Procedures	(2) Policies & Procedures	(3) Policies & Procedures
Emergency	-0.16**	-0.11	-0.12
	(0.05)	(0.06)	(0.06)
Alternate	-0.06	-0.05	-0.05
	(0.07)	(0.07)	(0.07)
Gender		0.16**	0.15**
		(0.05)	(0.06)
Black		-0.12	-0.13
		(0.06)	(0.06)
Hispanic Latino/a		0.04	0.04
		(0.10)	(0.10)
Other		0.01	0.02
		(0.08)	(0.08)
Degree Level		0.02	0.02
		(0.05)	(0.05)
High School			-0.00
			(0.06)
Middle School			-0.07
			(0.07)
Constant	3.77***	3.65***	3.67***
	(0.03)	(0.06)	(0.08)
Observations	793	793	793
Adjusted R-squared	0.01	0.02	0.02

Standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

In model 1, I introduced binary variables indicating certification status. Relative to standard certification status, emergency certification status is negatively associated with policies and procedures score ($p < .01$). The magnitude of the negative association is -0.16. Alternative certification status is not statistically associated with policies and procedures score when compared to standard certification status. In Model 2, teachers' sociodemographic background

characteristics are introduced for gender, race, and degree level. With the introduction of these variables, the negative association between certification status magnitude is -0.11 and is not statistically significant. Alternative certification status is not statistically associated with policies and procedures score after teachers' sociodemographic background factors are entered into the model. With regard to gender, being female is associated with higher policies and procedures scores than being male. The positive association is 0.16 and is statistically significant ($p < .001$). The additional background characteristics are not statistically associated with policies and procedures scores. In Model 3, controls for school level are added. Results remain largely similar across models. Relative to standard certification status, emergency certification is still negatively associated with policies and procedures score but is not statistically significant. The magnitude of the negative association is -0.12. Alternative certification status is not statistically associated with policies and procedures score when compared to standard certification status. The positive association of gender is 0.15 ($p < .001$). The sociodemographic information continues to have no statistically significant association with teachers' policies and procedures scores. Relative to elementary status, high school and middle school are not significantly associated with policies and procedures scores.

Chapter 5: Discussion

The purpose of this study was to explore the relationship between teachers' certification types and their quality, as measured through evaluation scores. This chapter discusses the key findings.

Summary of Key Findings

The results of the analysis showed consistent relationships between teachers' certification status and their evaluation scores. The results indicate that teachers who hold emergency certificates score at lower levels than teachers with standard certificates and that these differences are statically significant. The magnitude of the difference in scores is approximately 0.22 points on a scale that ranges from 0 to 4 using only whole number values for each scorable area. However, these estimates may be understated, as the robustness checks using ordinal regression indicate a magnitude of 0.80, which is almost one full change in the scale. The magnitude of the relationship between certification status and evaluation scores remained largely similar even after the introduction of sociodemographic background characteristics into the statistical models. All statistical analyses indicated a negative association between teachers who hold emergency certificates and final evaluation scores. The main results were consistent across a range of robustness checks.

In this section, I describe the findings of this study in greater detail. The analysis follows the hypotheses and tables presented in the previous chapter. The detailed findings firmly reject the idea that emergency-certified teachers have been entering the classroom to fill vacancies with high-quality teachers.

H^1: A difference exists in teacher quality between teachers who hold standard teacher certificates and those who hold alternative or emergency certificates.

As presented in the review of literature, few studies have used teacher certification scores to examine the quality of teachers and how that might be

associated with a teachers' certification type. Very few studies have investigated the relationship between certification type and the scores provided by a well thought out, data-driven, teacher evaluation tool. Instead, previous studies often include only students' academic performance on standardized assessments as a measure of teacher quality (Boyd, et.al, 2013; Goe, 2007; Wilson & Floden, 2003). However, the results of this study are still similar: teachers with emergency certificates continue to score lower even when a different measure is used. This research also supports the findings of Walsh & O'Tracy (2005) that the difference in benefits of a traditional standard certificate and an emergency certificate are small but present. Utilizing teacher evaluation scores as a means to measure teacher quality found consistently that teachers who hold emergency certificates scored lower on final evaluations than those who hold standard certificates. Emergency-certified teachers also scored lower in both subcategories *Critical Content* and *Rules and Procedures*. Controls were added throughout all the models to reject the assumption that additional variables, such as gender or ethnicity, played a part in a teachers' evaluation scores. The results of these models controlling for sociodemographic data continue to support the negative association between emergency certification and teacher evaluation scores.

H^2: A difference in teacher quality exists between new teachers who hold a standard teacher certificate and those who hold an emergency or alternative or emergency teaching certificate

The results for the tests of H^2, regarding new teachers align with the results of H^1 regarding all teachers. Emergency certification is significantly

negatively associated with evaluation scores, a measure of teacher quality, even when only new teachers are included in the sample. If new teachers have completed a teacher preparation program before entering the classroom, they have a higher total evaluation score and sub-scores in the areas of critical content and rules and procedures. These findings are similar the results of Boyd et al. (2008), who found that emergency-certified teachers with no experience have lower scores than those who hold a standard certification. My findings also mirror those of Walsh and O'Tracy (2007), who found that new teachers without preservice education may be less effective than those who do, but the difference is small. The only area in which emergency-certified teachers had no statistically significant negative association was within the subcategory indicator for policies and procedures. This indicates that emergency-certified teachers are able to follow policies and procedures; however, this is not an indicator of quality teaching in urban school settings as defined by the literature.

Empirical Literature

This study makes important contributions to the literature on teacher certification and teacher quality as measured by a teacher evaluation. Few studies have tested the relationship between teacher certification and teacher evaluation, a measure of quality. Darling-Hammond (2010) highlights the power of teacher education on student outcomes. Studies demonstrate that when a teacher has earned a teaching license and other credentials, has received a passing score in certification, and has taught for two or more years, the teacher is likely to have a large impact on student achievement. When teachers enter the field without

preparation or certification in the field, on the other hand, they have strong negative effects on student achievement, and the strongest negative effects are associated with inexperienced teachers working with temporary licenses (Darling-Hammond, 2010; Boyd, Lankford, Loeb, Rockoff, & Wyckoff, 2008). This study also supports the findings a 2008 study by the Louisiana Board of Regents that found teacher certification level to be a significant predictor of student achievement when using value-added methods. My study adds to the literature by reporting the statistically significant negative associations between an emergency certification and evaluation scores.

Teacher quality has traditionally been measured using VAM that rely on student assessment data (Boyd, et.al, 2013; Goe, 2007). These measures generally show that differences exist in teacher quality related to a teachers' certificate or years of teaching experience in most subject areas. Although these measures are easy for researchers to access, they fail to look at the whole teacher and fail to focus on the aspects of teaching that are under teachers' control. Goe (2007) points out that teacher quality is not one-dimensional. A variety of inputs and outputs of teaching might be appropriate measures depending on the situation. This study uses the MFTEM, which attempts to evaluate the whole teacher in a way that provides support for learning and growth. When used with fidelity, the tool helps critical teaching-related areas where teachers might need additional training or support. The teacher is then able to set goals and create a personal path to demonstrate mastery of an indicator. This study addresses the concerns of Goe

(2007) in that it attempts to use a teacher evaluation that encompasses many of the inputs and outputs of teacher quality.

The existing literature fails to highlight the teacher evaluation as a measure of teacher quality. Perhaps this gap has occurred because the tools used for teacher evaluation vary greatly from one state to the next, making it difficult to find generalizable trends. This research helps fill the gap on identifying teachers of quality by using MFTEM as a measure that directly reviews a teacher's level of proficiency in several key categories of quality teaching.

Limitations

One limitation of this study is that it investigated one large urban school district using one method for teacher evaluations. Furthermore, they study present non-causal results. The estimates offer suggestive evidence, but there may be other variables not modeled in this study's analyses that explain the relationship between certification status and evaluation score. A thorough analysis of the MFTEM indicates that it measures characteristics of a high quality, effective teacher, but other tools might result in different findings. This evaluation tool, while effective, is not reflective of every school district in Oklahoma or the United States. It measures different characteristics of teaching that might not be applicable to or reflective of all school districts. Future research could address these issues by selecting sites that would be able to evaluate teachers using multiple models to determine whether other models find the same association between certification type and teacher quality.

Second, the evaluators in the selected district preferred scoring in three categories, not all 23 contained within the MFTEM. The district did not insist that all teachers must be evaluated on all 23 scorable areas. It appears that the district instead asked evaluators to focus on three key areas during this school year. Therefore, the study was limited to only three sub-categories: (1) critical content, (2) rules and procedures, and (3) policy and procedures. These aspects of teaching, although important, are not the areas identified by the literature as indicators of high-quality teaching. It is not known why the school district avoided the indicators related to building relationships and strong communities, some of the most valuable attributes of a teacher in urban schools. Because so few indicators were included in the data, I was not able to test a third hypothesis that would have included identifying which indicator was the strongest predictor of a teacher's final evaluation score.

Third, not all teachers may have participated in the evaluation process. District policy states that career teachers who score a 3.5 or higher do not have to be evaluated every year. Not all teachers participated in the MFTEM during the 2018-19 school year. This may have also been due to a long-term illness keeping a teacher away from the classroom, or due to an evaluator's lack of participation in the process. The lack of complete teacher participation affected the sample size and may have resulted in a smaller magnitude of the negative associations identified in the study, because higher-scoring teachers were not included. Teachers who received an "innovative" score (the top score) during the previous school year were not required to participate in a teacher evaluation during 2018-

19. Lower participation in the evaluation process by higher scoring teachers may have reduced the association between certification and evaluation in this study. Future research controlling for these differences would advance this study and add to the literature.

An additional limit to this study is that evaluation scores were completed by a variety of administrators throughout several school sites. Although the district attempted to calibrate all evaluators, this study is unable to control for differences between observers' scores. Scores can be affected by human factors, such as an evaluator's concern that a low score might push a teacher out of the classroom. This human element should be considered when viewing the individual scores. The evaluation scores, although varied, are clustered around the mean, indicating that evaluators may have scored teachers similarly based on the classroom snapshot they observed. Bias and complacency are additional possibilities. If evaluators scored a teacher below level 3, they were required to write additional notes in the online portal, requiring additional work and time. This complacency may have led to an overestimate of some teachers' final evaluation scores to eliminate additional workload by the evaluator. While this study cannot determine whether bias or complacency played a direct role in this data set, future research might focus on this area to determine whether patterns are observed in teacher evaluation scores. Considering the limitations of this study, the results offer insightful evidence that should be interpreted cautiously.

Strengths

The strengths of this study come from the teacher evaluation model. Previous research indicates that the MFTEM can work for school districts that implement the program with fidelity and use it as a way to help teachers grow as educators. The model includes 23 key areas, all measured on a scale that can be easily defined by the users. The state of Oklahoma requires all school districts to conduct teacher evaluations, so data will continue to be collected for future research. The evaluators who implemented the evaluations participated in several calibration sessions where they learned how to use the MFTEM tool appropriately. This reinforces the validity of the dependent variables that were used throughout the research.

An additional strength of the study was that the evaluations were conducted before the COVID pandemic of 2020-2022, so the data contained evaluations of classes taught under normal circumstances. Data collected during the pandemic might not have been accurately assessed due to the continuous changes and adjustments to the school setting. As we are beginning a post-COVID school year in the fall of 2022, it would be interesting to see whether the data would show the same relationships.

If emergency-certified teachers are going to remain in the classroom, we should consider how they should be supported so that they are able to become more effective and achieve higher evaluation scores. Yearlong residencies with extended time in the field have been shown to improve teachers' value-added measures significantly (Darling-Hammond, 2010). Considering that one year spent observing and learning with a teacher would likely be a year without

adequate pay, this may not be an adequate solution for the short term. Instead, it may be important to consider the Teach for America (TFA) model. Although it is an alternative pathway, it has been shown to reduce the achievement gap between high-poverty and low-poverty schools in New York City (Boyd et al., 2008). TFA teachers attend summer training before entering the classroom and receive direct support from the directors though development sessions after contract hours. A month of summer preparation in a classroom with students may lead to higher evaluation scores and may meet the needs of community members who are interested in the teaching profession. Oklahoma currently requires that new and emergency-certified teachers be assigned a mentor, but personal experience indicates that it is not happening consistently because teachers are overworked and are not compensated for the additional mentoring workload. If the state focused more attention on supporting emergency teachers through strategic and financially supportive measures, the mentoring relationship may have an impact on emergency-certified teachers' evaluation scores. Guthery & Bailes (2022) found that new teachers who were provided supports such as mentorships were less likely to leave teaching and that teachers pursuing alternative paths to certification benefited most from mentorships (Smith & Ingersoll, 2004)

Policy makers have considered financial incentives for post-secondary students to enter the teaching profession. This includes a graduated scholarship and stipend for graduating and teaching in Oklahoma. As an example, the University of Oklahoma provides a TE plus option that allows students who are pursing arts and science degrees to add on a teaching degree program without

repeating content-area course work. TE plus stands for, Teacher Education: Professionalism, Leadership, Understanding, scholarship. These students become masters of the content along with learning the pedagogy and practices crucial for becoming successful teachers. Unfortunately, the outcomes will not become apparent for several years, as these students are currently enrolled in coursework.

Teacher shortages are continuing across the United States, and the number of education degrees awarded by colleges is falling (Garcia & Weiss, 2019b). Oklahoma continues to issue a large number of emergency certificates as a way to fill classroom vacancies (Palmer, 2021; Teacher Shortage Task Force, 2015), while promising that Emergency-certified teachers will be just as good as, or even better than, traditionally certified teachers because of the practical applications they have experienced while in the workplace. Although the magnitude was not large, the negative association found in this study between teachers' evaluation score and their certification type was statistically significant and adds to the literature on the quality of Emergency-certified teachers.

The state of Oklahoma has continued to reduce the requirements for teaching with a reduction in the number of exams teachers are required to take for certification (Teacher Shortage Task Force, 2015). The consequences of these practices can be seen in the quality of teaching, as measured by teacher evaluation scores, in the findings of this research. Emergency-certified teachers score lower on teacher evaluations than their peers who hold standard certificates. Supports should be put in place to reduce the difference in scores.

Areas for Future Research and Questions

Oklahoma has a teacher shortage, including in urban schools, which need the strongest teachers but are typically given the least qualified (Darling-Hammond, 2010; Lankford et. al, 2002). Oklahoma needs more teachers who can support student learning by building relationships, recognizing social justice, and working from a growth mind set. However, teachers who fill these critical roles often join the profession with no classroom experience due to fast-track entry (Chochran-Smith, 2004). Some Emergency-certified teachers begin the process of fulfilling requirements to receive a standard certificate through enrollment in a teacher preparation program. However, Walsh & Jacob (2007) found that even robust alternative certification programs, which include content aimed at improving teacher quality, often produce teachers whose quality ratings are not as high as those of traditionally certified teachers. Those with emergency certificates typically do not receive the kinds of training provided by alternative certification programs, much less traditional certification, so they have even less access to critical pedagogy training and fewer opportunities to improve their skills.

This quantitative study investigated the association between certification type and evaluation scores within one school district. However, it was not inclusive of all teachers, and there remains additional room for research. A particularly fertile area for future research would be through a longitudinal model, looking into sets of data over several years, to discover whether the results be similar each year or whether other patterns would arise through the analysis.

Another area of focus for future research should be centered on extending beyond this one urban school district. This study sought to address the gap in

literature because few studies on teacher quality use teacher evaluation as a measure. Although this study focuses on one school district, it would be informative to continue this line of investigation in other parts of Oklahoma, including rural school districts. Additionally, it would be fruitful to investigate other states' policies and procedures for emergency certification. An analysis of various states' practices and perspectives would help inform educational leaders and policy makers as they make decisions regarding the development of teachers who seek alternative pathways to the classroom.

This quantitative research addressed gaps in research surrounding teacher quality, teacher evaluation, and teacher certification. Although the research does not identify the precise teacher evaluation indicator that is most predictive of a teacher's evaluation score, it does identify an association between teachers' certification types and their overall final evaluation score. Policy regarding emergency certification of untrained teachers has a meaningful effect on Oklahoma's students. It is important for policy makers to re-evaluate the ease of entry into the teaching profession for those without foundational coursework in education. If those who hold a traditional teaching certificate outperform others on teacher evaluation, then one might wonder whether our communities should expect more from teachers with emergency certificates. Should they be asked to take a few introductory classes in the field of education before they walk into a classroom? Do policy makers understand the difference in teacher quality from one certification level to the next?

Policy makers can use this research to think about new ways of recruitment and how they might integrate key components of a standard certificate with the qualifications for an emergency certificate to address the lower teacher evaluation score. Policy makers may also consider a more holistic approach to evaluating teachers. This approach might include the students' feelings about the class, test scores, and feedback from an evaluation such as MFTEM.

Conclusion

This research is not generalizable to all school districts in Oklahoma, but it advances the literature on teacher evaluation and teacher certification. The lack of previous research in this area is significant, and this study opens the door for further research. This research took a quantitative approach to investigate the relationship between these aspects of teaching within Oklahoma. This study used pre-COVID, comprehensive data sets to study the relationship of a teachers' evaluation score, as a measure of quality, and a teachers' certification level.

Oklahoma issued 3,038 emergency teaching certificates for the 2018-19 school year and 3,321 for 2019-2020 (Palmer, 2021). Requests for emergency certificates have increased since the Oklahoma Task Force was created in 2015, and the trend indicates that requests will not return to the pre-2015 levels of less than 1000 per year (Palmer, 2015). This research study suggests that Oklahoma policy makers have not lived up to the promise of increasing the number of high-quality educators in the classroom by reducing barriers to entry. Instead, emergency-certified teachers score lower on teacher evaluations, a measure of quality, than their peers, and emergency-certified teachers are not staying in the

classroom to permanently fill teaching positions. In Texas, Guthery and Bailes note that those who enter the teaching profession using an alternate pathway are likely not to stay in the classroom, with most leaving within five years (Guthery & Bailes, 2022). Although this research provides no direct answer as to how Oklahoma might fix the teacher shortage, it does provide evidence that emergency-certified teachers score slightly lower than others. This is a concern because a student may not be able to make up for lost learning due to poor teaching. Students served by emergency-certified teachers may not acquire the level of proficiency required to achieve at the same level as their counterparts. If the number of emergency certificates issued continues to hold steady in Oklahoma, it raises concerns about the quality of education and our ability to compete academically against other states. Future research is needed to understand what adjustments Oklahoma might put in place to increase the quality of teaching provided by those who hold an emergency certificate. However, the findings of this research underscore the importance of teacher certification and its role in the quality of education throughout Oklahoma.

References

Aaronsohn, E., Carter, C. J., & Howell, M. (1995). Preparing monocultural teachers for a Multicultural world: Attitudes toward inner-city schools. *Equity and Excellence in Education*, *28*(1), 5-9.

Ashiedu, J. A., & Scott-Ladd, B. D. (2012). Understanding teacher attraction and retention drivers: Addressing teacher shortages. *Australian Journal of Teacher Education*, *37*(11), article 2.

Adams, C. M., & Forsyth, P. B. (2009). Conceptualizing and validating a measure of student Trust. In W. K. Hoy and M. DiPoala, M. (Eds.), *Studies in School Improvement* (pp. 263-277), Information Age.

Adams, C. M., & Forsyth, P. B. (2013). Revisiting the trust effect in urban elementary schools. *Elementary School Journal*, *114*(1), 1-21.

Allport, G. (1979). *The nature of prejudice* (25th anniversary ed.). Addison-Wesley.

Amin, J. N. (2016). Redefining the role of teachers in the digital era. *International Journal of Indian Psychology, 3*(3), 40-45.

Basileo, L. D., & Toth, M. (2019). A state level analysis of the Marzano Teacher Evaluation Model: Predicting teacher value-added measures with observation scores. *Practical Assessment, Research & Evaluation*, *24*(6), article 6.

Bowling, A. M., & Ball, A. L. (2018). Alternative certification: A solution or an alternative problem? *Journal of Agricultural Education, 59*(2), 109-122.

Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2013). Analyzing the determinants of the matching of public-school teachers to jobs: Disentangling the preferences of teachers and employers. *Journal of Labor Economics*, *31*(1), 83-117.

Boyd, D., Grossman, P., Hammerness, K., Lankford, H., Loeb, S., Ronfeldt, M., & Wyckoff, J. (2012). Recruiting effective math teachers: Evidence from New York City. *American Educational Research Journal*, *49*(6), 1008-1047.

Boyd, D., Lankford, H., Loeb, S. Rockoff, J., & Whycoff, J. (2008). The narrowing gap in New York City teacher qualifications and its implications for student achievement in high-poverty schools. National Center for the Analysis of Longitudinal Data in Education Research.

Burkhardt, J. (2017). Revamped Oklahoma program recruits high school students to be teachers. American Board. https://www.americanboard.org/blog

Certo, J. L., & Fox, J. E. (2002). Retaining quality teachers. *High School Journal, 86*(1), 57-75.

Chapman, T. K. (2013). You can't erase race! Using CRT to explain the presence of race and racism in majority white suburban schools. *Discourse: Studies in the Cultural Politics of Education, 34*(4), 611-627.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers. I: Evaluating bias in teacher value-added estimates. *American Economic Review, 104*(9), 2593-2632.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers. II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633-79.

Chism, N. V. N. (1999). *Peer review of teaching. A sourcebook*. Anker Publishing.

Cochran-Smith, M. (2004). Stayers, leavers, lovers, and dreamers: Insights about teacher retention. *Journal of Teacher Education*, *55*(5), 387-392.

Cochran-Smith, M. (2003). Teaching quality matters. *Journal of Teacher Education, 54*(2), 95-98.

Cohen, J. (1992). A power primer. *Quantitative Methods in Psychology*, *112*(1), 155-159.

Connally, K., & Tooley, M. (2016). Beyond ratings: Re-envisioning state teacher evaluation systems as tools for professional growth. *New America, 4*(2), 131-144.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage.

Cullen, T. (2018, April 4). Teacher exodus update. http://tacullen.oucreate.com/

Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan, 93*(6), 8-15.

Darling-Hammond, L. (2010). Teacher education and the American future. *Journal of Teacher Education*, *61*(1-2), 35-47.

Darling-Hammond, L., Berry, B., & Thoreson, A. (2001). Does teacher certification matter? Evaluating the evidence. *Educational Evaluation and Policy Analysis*, *23*(1), 57-77.

Darling-Hammond, L., & Youngs, P. (2002). Defining "highly qualified teachers": What does "scientifically-based research" actually tell us? *Educational Researcher*, *31*(9), 13-25.

DiBara, J. A. (2007). Responsible to the kids: The goals and struggles of urban high school teachers. *New Educator,* 3(1), 11-30.

Donaldson, M. L. (2009). Into and out of city schools: The retention of teachers prepared for urban settings. *Equity & Excellence in Education*, *42*(3), 347-370.

Donnelly, R. (2007). Perceived impact of peer observation of teaching in higher education. *International Journal of Teaching and Learning in Higher Education*, *19*(2), 117-129.

Duckworth, A. L., Quinn, P. D., & Seligman, M. E. (2009). Positive predictors of teacher effectiveness. *Journal of Positive Psychology, 4*(6), 540-547.

Duncan-Andrade, J. (2007). Gangstas, wankstas, and ridas: Defining, developing, and supporting effective teachers in urban schools. *International Journal of Qualitative Studies in Education, 20*(6), 617-630.

Every Student Succeeds Act, 20 U.S.C. § 6301 (2015). https://www.congress.gov/bill/114th-congress/senate-bill/1177

Feuer, M. J., Floden, R. E., Chudowsky, N., & Ahn, J. (2013). Evaluation of teacher preparation programs: Purposes, methods, and policy options. National Academy of Education.

Fine, R. (2019). Beyond teacher pay: Class size matters. Oklahoma Policy Institute. https://okpolicy.org/beyond-teacher-pay-class- size-matters/.

Ford, T. G., Urick, A., & Wilson, A. (2018). Exploring the effect of supportive teacher evaluation experiences on U.S. teachers' job satisfaction. *Educational Policy Analysis Archives, 26*(59), 1-36.

Ford, T. G., Van Sickle, M. E., Clark, L. V., Fazio-Brunson, M., & Schween, D. C. (2017). Teacher self-efficacy, professional commitment, and high-stakes teacher evaluation policy in Louisiana. *Educational Policy*, *31*(2), 202-248.

Forsyth, P. B., Adams, C. M., & Barnes, L. (2002). Parental trust of school: Scale development [paper presentation]. American Educational Research Association, Division L., New Orleans, LA, United States.

Fraser, B. J. (2018). Milestones in the evolution of the learning environments field over the past three decades. In D. Zandvliet and B. J. Fraser (Eds.), *Thirty years of learning environments: Looking back and looking forward* (pp. 1-19). Brill.

Fraser, J. W. (2001). *The school in the United States: A documentary history*. McGraw Hill.

Freedman, S. W., & Appleman, D. (2009). "In it for the long haul": How teacher education can contribute to teacher retention in high-poverty, urban schools. *Journal of Teacher Education*, *60*(3), 323-337.

Freedman, S. W., & Appleman, D. (2008). "What else would I be doing?" Teacher identity and teacher retention in urban schools. *Teacher Education Quarterly*, *35*(3), 109-126.

Friesen, S., & Jardine, D. (2009). 21st century learning and learners. A report prepared for Western and Northern Canadian Curriculum Protocol by Galileo Educational Network.

Fenstermacher, G., & Richardson, V. (2005). On making determinations of quality in teaching. *Teachers College Record, 107*(1), 186-213.

Garcia, E., & Weiss, E. (2019a). The teacher shortage is real, large and growing, and worse than we thought. The first report in "The Perfect Storm in the Teacher Labor Market" series. Economic Policy Institute.

Garcia, E., & Weiss, E. (2019b). U.S. schools struggle to hire and retain teachers. The second report in "The Perfect Storm in the Teacher Labor Market" series. Economic Policy Institute.

Gardner, A., & Brindis, C. (2017). Advocacy and policy change evaluation: Theory and practice. Stanford University Press.

Gasser, L., Gutter, J., Buholzer, A., & Wettstein, A. (2018). Emotionally supportive classroom interactions and students' perceptions of their teachers as caring and just. *Learning and Instruction, 54,* 82-92.

Gibbs, G. (1995). How can promoting excellent teachers promote excellent teaching? *Innovations in Education and Training International*, *32*(1), 74-81.

Goddard, R. D., & Goddard, Y. L. (2001). A multilevel analysis of the relationship between teacher and collective efficacy in urban schools. *Teaching and Teacher Education*, *17*(7), 807-818.

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

Goe, L. (2007). The link between teacher quality and student outcomes: A research synthesis. National Comprehensive Center for Teacher Quality.

Goe, L., Biggers, K., & Croft, A. (2012). Linking teacher evaluation to professional development: Focusing on improving teaching and learning. Research & policy brief. National Comprehensive Center for Teacher Quality.

Goldhaber, D. (2019). Evidence-based teacher preparation: Policy context and what we know. *Journal of Teacher Education, 70*(2), 90-101.

Gordon, R. J., Kane, T. J., & Staiger, D. (2006). *Identifying effective teachers using performance on the job*. Brookings Institution.

Gorski, P. C. (2016). Poverty and the ideological imperative: A call to unhook from deficit and grit ideology and to strive for structural ideology in teacher education. *Journal of Education for Teaching*, *42*(4), 378-386.

Grasha, A. F. (1994). A matter of style: The teacher as expert, formal authority, personal model, facilitator, and delegator. *College Teaching, 42*(4), 142-149.

Guin, K. (2004). Chronic teacher turnover in urban elementary schools. *Education Policy Analysis Archives*, *12*, 42.

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

Halverson, R., Kelly, C., & Kimball, S. (2004). Implementing teacher evaluation systems: How principals make sense of complex artifacts to shape local instructional practice. *Educational Administration, Policy, and Reform: Research and Measurement*, 3, 153-188.

Hamlin, D. (2021). Can a positive school climate promote student attendance? Evidence from New York City. *American Educational Research Journal*, *58*(2), 315-342.

Hamlin, D., & Li, A. (2020). The relationship between parent volunteering in school and school safety in disadvantaged urban neighborhoods. *Journal of School Violence*, *19*(3), 362-376.

Hamre, B. K., & Pianta, R. C. (2001). Early teacher-child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, *72*(2), 625-638.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, *30*(3), 466-479.

Haviland, V. S. (2008). "Things get glossed over": Rearticulating the silencing power of whiteness in education. *Journal of Teacher Education*, *59*(1), 40-54.

Hazi, H. M. (2014). The marketing of teacher evaluation: The seductive claims of instruments. *WERA Educational Journal*, *6*(1), 2-9.

Hollins, E. R. (2011). Teacher preparation for quality teaching. *Journal of Teacher Education*, *62*(4), 395-407.

Horng, E.L., Klasik, D., & Lobe, S. (2010). Principal time use and school effectiveness. *American Journal of Education, 116,* 491-523.

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

Huang, F. L., Lewis, C., Cohen, D. R., Prewett S., & Herman, K. (2018). Bullying involvement, teacher-student relationships, and psychosocial outcomes. *School Psychology Quarterly, 33*(2), 223.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal*, *38*(3), 499-534.

Ingersoll, R. M. (2002). The teacher shortage: A case of wrong diagnosis and wrong prescription. *NASSP Bulletin*, *86*(631), 16-31.

Ingersoll, R. (2004). Four myths about America's teacher quality problem. In M. A. Smylie & D. Miretzky (Eds.),, *Developing the Teacher Workforce, 103rd Yearbook of the National Society for the Study of Education (pp. 1-33)*. University of Chicago Press.

Ingersoll, R. M. (2012). Beginning teacher induction what the data tell us. *Phi Delta Kappan*, *93*(8), 47-51.

Ingersoll, R. M., & Smith, T. M. (2003). The wrong solution to the teacher shortage. *Educational Leadership*, *60*(8), 30-33.

Jacob, B. A. (2007). The challenges of staffing urban schools with effective teachers. *The Future of* Children, 17(1), 129-153.

Jones, J. R., & Henry, M. (2019). Students' assessment of teacher quality: recommendations for improvement [paper presentation]. Southern Regional Council on Educational Administration, Fort Myers, FL, United States.

Joyce, J., Gitomer, D., & Iaconangelo, C. (2018). Classroom assignments as measures of teaching quality. *Learning and Instruction*, 54, 48-61.

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

Kane, T. J., & Staiger, D. O. (2012). Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains. Research paper. MET Project. Bill & Melinda Gates Foundation.

Kamenetz, A. (2014). For teachers, many paths into the classroom … Some say too many. *NPREd.* https://www.npr.org/sections/ed/2014/09/12/347375798/for-teachers-many-paths-into-the-classroom-some-say-too-many.

Keller, B. (2007). NCLB rules on "quality" fall short. *Education Week*, *26*(37), 1-16.

Kirby, S. N., Darling-Hammond, L., & Hudson, L. (1989). Nontraditional recruits to mathematics and science teaching. *Educational Evaluation and Policy Analysis*, *11*(3), 301-323.

Killion, J. (2015). High-quality collaboration benefits teachers and students. *Learning Professional, 36*(5), 62.

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

Laczko-Kerr, I., & Berliner, D. C. (2002). The effectiveness of "Teach for America" and other under-certified teachers. *Education Policy Analysis Archives, 10*(37).

Ladson-Billings, G. J. (1999). Preparing teachers for diverse student populations: A critical race theory perspective. *Review of Research in Education, 24*(1), 211-247.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, *24*(1), 37-62.

Lazarte, N. (2018). *2018 Oklahoma educator supply & demand report: Trends, projections, and recommendations.* Educator Effectiveness & Policy Research, Oklahoma State Department of Education.

Learning Policy Institute (2018). Uncertified teachers and teacher vacancies by state. https://learningpolicyinstitute.org/uncertified-teachers-and-teacher-vacancies-state

Leland, C. H., & Harste, J. C. (2005). Doing what we want to become: Preparing new urban teachers. *Urban Education*, *40*(1), 60-77.

Lempesis, C. (1984). Peer observation improves teacher performance. *NASSP Bulletin*, *68*(471), 155-156.

Lewis, R., & Lovegrove, M. N. (1987). The teacher as disciplinarian: How do students feel? *Australian Journal of Education, 31*(2), 173-186.

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.

Loeb, S., Darling-Hammond, L., & Luczak, J. (2005). How teaching conditions predict teacher turnover in California schools. *Peabody Journal of Education*, *80*(3), 44-70.

Louisiana Board of Regents. (2008). Regents study shows teacher certification matters. https://regents.state.la.us/assets/docs/2013/09/StudyShowsTchrCertMatters-12.3.08.pdf

Lowenstein, K. L. (2009). The work of multicultural teacher education: Reconceptualizing white teacher candidates as learners. *Review of Educational Research*, *79*(1), 163-196.

Maiden, J., & Reynolds, T. (2019). Oklahoma. *Journal of Education Finance*, *44*(3), 318-320.

Marsh, J. A., Springer, M. G., McCaffrey, D. F., Yuan, K., & Epstein, S. (2011). *A big apple for educators: New York City's experiment with schoolwide performance bonuses: Final evaluation report.* Rand Corporation.

Marzano, R. J., & Toth, M.D. (2017). The Marzano focused teacher evaluation model: A focused, scientific-behavioral evaluation model for standards-based classrooms. https://www.marzanocenter.com

Marzano, R., Toth, M., & Schooling, P. (2012). Examining the role of teacher evaluation in student achievement: Contemporary research base for the Marzano Causal Teacher Evaluation Model. Marzano Center. http://www.marzanocenter.com/files/MC_White_Paper_20120424.Pdf

Maslach, C., & Leither, M.P. (2008). *The truth about burnout: How organizations cause personal stress and what to do about it.* John Wiley & Sons.

Mobra, T., & Hamlin, D. (2020). Emergency-certified teachers' motivations for entering the teaching profession: Evidence from Oklahoma. *Education Policy Analysis Archives*, *28*(109).

National Commission on Teaching and America's Future. (2003). No dream denied: A pledge to America's children. Summary report. https://www.alaska.edu/asmp/publications/nct_NoDreamDenied_2003.pdf

National Research Council. (2010). Preparing teachers: Building evidence for sound policy. National Academies Press.

No Child Left Behind Act of 2001, P.L.107-110, 20 U.S.C. § 6319 (2002).

Noell, G. H., Burns, J. M., Gansle, K. A. (2019). Linking student achievement to teacher preparation: emergent challenges in implementing value added assessment. *Journal of Teacher Education, 70*(2), 128-138.

Olsen, B., & Anderson, L. (2007). Courses of action: A qualitative investigation into urban teacher retention and career development. *Urban Education, 42*(1), 5-29

Oklahoma City Public Schools. (n.d.) Teacher and leader effectiveness district teacher guidelines 2018-2019. https://www.nctq.org/dmsView/TLE_Guidelines_2018_2019

Oklahoma Education Coalition. (2021). The facts about Oklahoma education. https://www.okedcoalition.org/get-the-facts.html

Oklahoma State Courts Network. (2012). Question submitted by the honorable Ed Cannaday, State Representative, District 15. http://www.oscn.net/applications/oscn/DeliverDocument.asp?CiteID=465651

Oklahoma State Department of Education (2022). Alternative paths to teacher certification. https://sde.ok.gov/alternative-paths-teacher-certification

Oklahoma State Department of Education (2021). Emergency certification guidance. https://sde.ok.gov/emergency-certification- guidance

Oklahoma State Department of Education (2019). Educator effectiveness theory of action. http://ok.gov/sde/tle.

Oklahoma State Department of Education (2018). State Board of Education–approved emergency certification applications. https://sde.ok.gov/sites/default/files/documents/files/Emergency%20Certifications.pdf

Oklahoma State Department of Education. (2019). Oklahoma alternative placement program for teacher certification. https://sde.ok.gov/oklahoma-alternative-placement-program-teacher- certification

Oklahoma State Department of Education. (2018). Teacher Shortage Task Force identifies progress in combating statewide crisis. https://sde.ok.gov/newsblog/2018-09-13/teacher-shortage-task-force-identifies-progress-combating-statewide-crisis

Oklahoma State School Boards Association. (2019). Survey: Morale, pay and teacher numbers rising but shortage remains. https://www.ossba.org/2019/08/20/survey-morale-pay-and-teacher-numbers-rising-but-shortage-remains/

Oklahoma State School Boards Association. (2017). District leaders describe mixed experiences with emergency-certified teachers. https://www.ossba.org/2017/08/22/oklahomas-teacher-shortage-deepens/

Oklahoma Teacher and Leader Effectiveness Evaluation System, 70 OK Stat § 70-6-101.16.

Palmer, Jennifer. (2021, September 1). Oklahoma starts year with its most emergency teachers on record. *Oklahoma Watch*. https://oklahomawatch.org/2021/09/01/oklahoma-starts-year-with-most-emergency-teachers-on-record/

Palmer, Jennifer. (2019, February 1). State leaders face dilemma in addressing class sizes. *Oklahoma Watch.* https://oklahomawatch.org /2019/02/01/state-leaders-face-dilemma-in-addressing-class-sizes/

Patton, L. D., & Jordan, J. L. (2017). It's not about you, it's about us: A Black woman administrator's efforts to disrupt white fragility in an urban school. *Journal of Cases in Educational Leadership*, *20*(1), 80-91.

Pressley, T., Roehrig, A. D., & Turner, J. E. (2018). Elementary teachers' perceptions of a reformed teacher-evaluation system. *Teacher Educator*, *53*(1), 21-43.

Pollock, M., Bocala, C., Deckman, S. L., & Dickstein-Staub, S. (2016). Caricature and hyperbole in preservice teacher professional development for diversity. *Urban Education*, *51*(6), 629-658.

Qu, Y., & Becker, B. J. (2003). Does traditional teacher certification imply quality? A meta-analysis [paper presentation]. Annual Meeting of the American Educational Research Association, Chicago, IL, United States.

Rahman, T., Fox, M. A., Ikoma, S., & Gray, L. (2017). Certification status and experience of U.S. public school teachers: Variations across student subgroups. NCES 2017-056. National Center for Education Statistics.

Ravid, R. (2019). *Practical statistics for educators*. Rowman & Littlefield.

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

Redding, C., & Nguyen, T. D. (2020). Recent trends in the characteristics of new teachers, the schools in which they teach, and their turnover rates. *Teachers College Record, 122*(7), 1-36.

Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes*. Economic Policy Institute.

Rice-Boothe, M. B. (2001 ,September 9). Students want to know: Will I feel safe, respected and valued? https://www.leadershipacademy.org/blog/students-want-to-know-will-l-feel-safe-respected-and-valued/

Roorda, D. L., Koomen, H. M., Spilt, J. L., & Oort, F. J. (2011). The influence of affective teacher-student relationships on students' school engagement and achievement: A meta-analytic approach. *Review of Educational Research*, *81*(4), 493-529.

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, *50*(1), 4-36.

Sabine, G.A. (1971). *How students rate their schools and teachers.* National Association of Secondary School Principals

Sachs, S. K. (2004). Evaluation of teacher attributes as predictors of success in urban schools. *Journal of Teacher Education, 55*(2), 177-187.

Scribner, J. D., & Layton, D. H. (Eds.). (2003). *The study of educational politics: The 1994 commemorative yearbook of the politics of education association 1969-1994.* Routledge.

Shevalier, R., & McKenzie, B.A. (2012). Culturally responsive teaching as an ethics- and care-based approach to urban education. *Urban Education*, *47*(6), 1086-1105.

Shen, J., Leslie, J., Spybrook, J., & Ma, X. (2012). Are principal background and school processes related to teacher job satisfaction? A multilevel study using schools and staffing survey 2003-04. *American Educational Research Journal, 49*, 200-230.

Shultz, E., Neyhart, K., & Reck, U. M. (1996). Swimming against the tide: A study of prospective teachers' attitudes regarding cultural diversity and urban teaching. *Western Journal of Black Studies*, *20*(1), 4-29.

Sleeter, C. E. (2017). Critical race theory and the whiteness of teacher education. *Urban Education, 52*(2), 155-169.

Sleeter, C. E. (2001). Preparing teachers for culturally diverse schools research and the overwhelming presence of whiteness. *Journal of Teacher Education*, *52*(2), 94-106.

Smith, T. M., Ingersoll R. M. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Educational Research Journal, 41*, 681-714.

Solórzano, D. G. (1997). Images and words that wound: Critical race theory, racial stereotyping, and teacher education. *Teacher Education Quarterly, 24*(3), 5-19.

Steele, J. L., Pepper, M. J., Springer, M. G., & Lockwood, J. R. (2015). The distribution and mobility of effective teachers: Evidence from a large, urban school district. *Economics of Education Review*, *48*, 86-101.

Stinberg, M. P., & Sartain, L. (2015). Does teacher evaluation improve school performance? Experimental evidence from Chicago's excellence in teaching project. *Journal of Policy Analysis and Management, 10,* 535-572.

Stronge, J. H., Ward, T. J., Tucker, P. D., & Hindman, J. L. (2007). What is the relationship between teacher quality and student achievement? An exploratory study. *Journal of Personnel Evaluation in Education*, *20*(3-4), 165-184.

Sutcher, L., Darling-Hammond, L., & Carver-Thomas, D. (2016). A coming crisis in teaching? Teacher supply, demand, and shortages in the U.S. Research brief. Learning Policy Institute.

Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review, 102*(7), 3628-51.

Teach for America. (2022). Our program. https://www.teachforamerica.org/experience/our-program

Teacher Shortage Task Force. (2015). *Preliminary report December 2015.* Report prepared for Oklahoma State Department of Education.

Teacher Shortage Task Force. (2018). *August 2018 update.* Report prepared for Oklahoma State Department of Education.

Tourkin, S. C., Pugh, K. W., Fondelier, S. E., Parmer, R. J., Cole, C., Jacson, B., Warner, T., Weant, G., and Walter, E. *1999-2000 schools and staffing survey (SASS) data file user's manual*. NCES 2004-303. National Center for Education Statistics. U.S. Department of Education

Tulsa Public Schools. (2015). *TLE observation & evaluation handbook for evaluators using the Tulsa Model.* https://www.nctq.org/dmsView/Handbook_TLE_Observation_and_Evaluation_System_74031

Tulsa Public Schools. (2022). Evaluation resources. https://www.tulsaschools.org/teamtulsa

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

U.S. Department of Education. (2002). *Meeting the highly qualified teachers challenge: The Secretary's annual report on teacher quality.* Office of Postsecondary Education.

Van Cauwenberg, G. (2010, October 12). *Effective vs. efficient when it comes to defining quality.* https://www.opexsociety.org.

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science research. *European Journal of Information Systems*, *25*(1), 77-89.

Von Hippel, P. T., & Bellows, L. (2018). How much does teacher quality vary across teacher preparation programs? Reanalyses from six states. *Economics of Education Review*, *64*, 298-312.

Walsh, K., & Jacobs, S. (2007). Alternative certification isn't alternative. Thomas B. Fordham Institute.

Walsh, K., & O'Tracy C. (2005). *Increasing the odds: How better policies can yield good teachers.* National Center for Teacher Quality. https://www.nctq.org/nctq/images/nctq_io.pdf

What is evaluation? (n.d.). https://www.informalscience.org/what-evaluation-0

Wilkins, J. (2014). Good teacher-student relationships: Perspectives of teachers in urban high schools. *American Secondary Education, 43*(1)52-68.

Wilson, S. M., & Floden, R. E. (2003). *Creating effective teachers: Concise answers for hard questions. An addendum to the report "Teacher preparation research: Current knowledge, gaps, and recommendations."* AACTE Publications.

Wilson, S. M., Floden, R. E., & Ferini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations.* A research report prepared for the U.S. Department of Education by the Center for the Study of Teaching and Policy in collaboration with Michigan State University.