INVESTIGATING RELATIONSHIPS BETWEEN MEASURES OF TEACHER QUALITY AND SCHOOL CONTEXT IN COLORADO: A MULTILEVEL MODERATED MEDIATION ANALYSIS

by

CURTIS LEE GARCIA

B.A., Yale University, 2008

M.A., Adams State University, 2011

A dissertation submitted to the Graduate Faculty of the University of Colorado Colorado Springs in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Leadership, Research, and Foundations

2020
This dissertation for the Doctor of Philosophy degree by

Curtis Lee Garcia

has been approved for the

Department of Leadership, Research, and Foundations

by

Robert Mitchell, Chair

Joseph Taylor

Sylvia Mendez

Grant Clayton

Stephanie Hensley

Date _______________
Garcia, Curtis Lee (Ph.D., Educational Leadership, Research and Policy)
Investigating Relationships between Measures of Teacher Quality and School Context in Colorado: A Multilevel Moderated Mediation Analysis

Dissertation directed by Assistant Professor Robert Mitchell

ABSTRACT

Recent education reform efforts have used high stakes teacher evaluations that include measures of student performance as a strategy for enhancing teacher quality and student achievement. While research has raised some doubt about the effectiveness of such policies to achieve their goals, there remains a need to understand how various measures of teacher quality relate to one another and the extent to which school contexts affect measures of teacher quality. This study utilized a quantitative multilevel moderated mediation analysis to investigate the relationships between measures of teacher quality based on qualifications, those based on evaluations of teacher practice, and those based on teacher effects on student outcomes. The findings of the analysis revealed that teacher education level and years of experience are positively associated with higher teacher effect estimates and these effects are partially mediated by teacher practices. The rural status of the school was found to have no significant effect on the relationships between teacher qualifications, teacher practices, and teacher effect estimates. The findings of this study contribute to theoretical understandings of teacher quality and provide guidance to policy makers about how constructs of teacher quality and school context might be applied within large-scale teacher evaluation policy frameworks.
DEDICATION

This dissertation is dedicated to my loving family who has provided me with endless love, patience, and support throughout my life. I would like to acknowledge my life-partner, Leo Martinez, who stood by my side throughout every step of my doctoral degree and gave me love and encouragement when I needed it most. His belief in me kept me going when the work became tough and my stress levels got the best of me. I cannot thank him enough for the love and care he gave me during all the sleepless nights, the travels back and forth between Alamosa and Colorado Springs, and all the times I needed reminders that “everything will be OK.” I love you, my Leo!

My parents, Lee and Cordy Garcia, along with my sister, Kimberly Garcia, have always inspired me to achieve my goals and have provided me with endless love and support throughout my life. It is because of their unconditional love that I have managed to achieve success in my education and career. There is no happiness that can compare to that of being with family and I owe all I have to Mom, Dad, and Kim. I love you guys!

This research is also dedicated to my Padrino, Felix Garcia, and my aunt, Patsy Garcia, who graciously opened their home to me while I completed my doctoral studies. Without their love, support, and amazing hospitality, I would not have been able to sustain the two years of long, late-night trips between Alamosa and Colorado Springs while I completed my degree. I love you both and thank you for all the support.

Finally, this dissertation is dedicated to all of the hard working teachers and leaders whom I have had the privilege of working with throughout my career as an educator. It is their passion and dedication for helping all kids that inspires me in my work as an educational leader.
ACKNOWLEDGEMENTS

I wish to acknowledge Dr. Robert Mitchell for serving as my Chair during my dissertation. His words of encouragement and guidance throughout this research enabled me to get through challenging times and reminded me about the value of caring educators during any stage of one’s educational career. My hopes are that we continue to work as colleagues within the field of rural education for years to come. I would also like to acknowledge Dr. Joseph Taylor, who served as my methodologist. His patience and kindness, along with his superior knowledge of quantitative methods, were invaluable to me as I completed this research. Dr. Sylvia Mendez, who served as my faculty advisor and member of my dissertation committee, provided ongoing support during my time in the program and I am forever grateful for her kindness and support. I would also like to thank Dr. Grant Clayton for serving as a member of my committee and providing me with feedback that pushed my thinking as I completed my research.

A special acknowledgement is extended to my dear friend and former colleague, Dr. Stephanie Hensley, who not only served as a member of my committee but has also helped me in so many ways within my career in higher education. She is truly an inspirational educator and I continue to look up to her for all the amazing work she does for kids and teachers.

Finally, I would like to thank all of my colleagues and friends at Adams State University and the Colorado Department of Education who inspired me throughout this process. Thank you, Chrissy McKinney, Angela Madrid, Dr. Cheri Meder, Mary Bivens, Courtney Cabrera, Sue Gill, Lynn Kintz, Karin Reynolds, Joslyn Robich, Rachel Paul, Carolyn Haug, and Molly Gold for all of your support.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Problem Statement</td>
<td>3</td>
</tr>
<tr>
<td>Purpose Statement</td>
<td>5</td>
</tr>
<tr>
<td>Research Questions</td>
<td>7</td>
</tr>
<tr>
<td>Overview of Methodology</td>
<td>9</td>
</tr>
<tr>
<td>Overview of Teacher Evaluation in Colorado</td>
<td>10</td>
</tr>
<tr>
<td>SB 10-191 Design and Goals</td>
<td>11</td>
</tr>
<tr>
<td>Teacher Quality Standards</td>
<td>12</td>
</tr>
<tr>
<td>Measures of Student Learning</td>
<td>14</td>
</tr>
<tr>
<td>Distribution of Teacher Quality Under SB 10-191</td>
<td>15</td>
</tr>
<tr>
<td>Conceptual and Theoretical Frameworks</td>
<td>16</td>
</tr>
<tr>
<td>Dynamic Theory of Educational Effectiveness</td>
<td>17</td>
</tr>
<tr>
<td>Teacher Quality Frameworks</td>
<td>18</td>
</tr>
<tr>
<td>Conceptualizing School Context</td>
<td>19</td>
</tr>
<tr>
<td>Application of Theoretical Models and Hypotheses</td>
<td>20</td>
</tr>
<tr>
<td>Rationale and Significance</td>
<td>22</td>
</tr>
<tr>
<td>Organization of the Dissertation</td>
<td>23</td>
</tr>
<tr>
<td>II. LITERATURE REVIEW</td>
<td>25</td>
</tr>
<tr>
<td>Historical Context: Issues of Teacher Quality in Education Reform</td>
<td>26</td>
</tr>
<tr>
<td>Measuring Teacher Quality through Evaluation</td>
<td>28</td>
</tr>
<tr>
<td>Teacher Quality</td>
<td>31</td>
</tr>
</tbody>
</table>
LIST OF TABLES

TABLE

1. Descriptive Statistics of Research Sample.............................................................. 71

2. Cross Tabulation of Teacher Education Level and Years of Experience .............. 72

3. Parameter Estimates for Ordinal Logistic Regression of SB 10-191 Teacher Effect
   Estimates on Teacher Qualifications ....................................................................... 74

4. Parameter Estimates for Mediation Analysis (a paths).............................................. 76

5. Parameter Estimates for Mediation Analysis (b and c’ paths)................................. 77

6. Indirect Effects of Teacher Qualifications on SB 10-191 Teacher Effect Estimates . 80

7. Interaction Effects of Rural Status on Teacher Qualifications, SB 10-191 Professional
   Practices, and SB 10-191 Teacher Effect Estimates............................................... 81
LIST OF FIGURES

FIGURE

1. Moderated Mediation Analysis of Teacher Qualifications, School Context, and SB 10-191 Evaluations Ratings ................................................................. 7

2. Composition of overall teacher evaluation ratings based on SB 10-191 .............. 12

3. Conceptual Framework developed for this study using a dynamic model of educational effectiveness. ........................................................................................... 21


5. Analytical Model for Moderated Mediation of Teacher Qualifications, Teacher Practices, and Teacher Effect Estimates ............................................ 51


7. Statistical Model for Moderated Mediation of Teacher Qualifications, Teacher Practices, and Teacher Effect Estimates with Variable labels .................. 73
CHAPTER I
INTRODUCTION

In 2010, the state of Colorado launched its first statewide evaluation system for teachers and principals with the passage of the Great Teachers and Leaders Law (Colorado Senate, 2010). Commonly referred to as SB 10-191, the law became part of Colorado’s three-pronged approach to school reform under President Barack Obama’s Race to the Top (RttT, 2011) Initiative. The reform initiative also included new academic standards and assessments aligned to the Common Core State Standards, thus forming the basis for Colorado’s broad strategy to improve student and teacher performance (Colorado Department of Education (CDE), 2018a).

In an effort to build upon previous attempts to enhance educational quality in Colorado schools, SB 10-191 established several new requirements for educator evaluations. This included, but was not limited to, a mandate that all teachers and principals receive annual evaluations based upon a common set of quality standards that include the results of student assessments. In addition, the law made significant changes to Colorado’s teacher tenure process by making tenure contingent upon three consecutive years of positive evaluation ratings and providing for the removal of tenure if teachers fail to maintain expected levels of performance as determined by proficiency targets as established by the State Board of Education (Colorado Senate, 2010).

Since its inception, SB 10-191 has sparked heated debates among education scholars, policy makers, and practitioners about the use of teacher evaluation as a policy strategy to enhance teaching and learning. These debates have largely focused on issues
related to the purpose of evaluation (Almy, 2011; Curtis & Wiener, 2012; Donaldson & Papay, 2015; Fullan, 2011; Hallinger et al., 2014; Hanushek, 2009; Mehta & Fine, 2015; Papay, 2012; Thomas et al., 2010), the design of evaluation systems (Bill and Melinda Gates Foundation, 2013; Doherty & Jacobs, 2015; Education Commission of the States, 2018; Frontier & Mielke, 2016), implementation of evaluation policies and systems (Grissom & Loeb, 2017; Harris & Sass, 2014; Jacob & Lefgren, 2008; Lavigne, 2018; Ramirez et al., 2014; Simpson, 2014) and measuring the effects evaluation policies on teacher satisfaction (Koedel et al., 2017), the teacher labor market (Kraft et al., 2018), and student achievement (Farley & Chamberlain, 2018; Kraft et al., 2018; Kraft & Gilmour, 2017; Schoales, 2017; Stecher et al., 2018). Proponents of educator evaluation laws point to findings, which suggest that regular observation and evaluation of educators, combined with high-quality feedback, can support improvements in student outcomes by helping to identify effective teaching practices, while measuring professional growth in ways that contribute to expertise and better teaching (Bill and Melinda Gates Foundation, 2013; Frontier & Mielke, 2016). Critics have argued that large-scale teacher evaluation policies have suffered from poor implementation (Ramirez et al., 2014), have failed to produce distributions in evaluation ratings that look different than before the policies were passed (Kraft & Gilmour, 2017; Schoales, 2017), and have failed to improve student outcomes or close achievement gaps among low-income and minority students (Stecher et al., 2018).

In spite of the mixed findings of research on educator evaluation policies (Putman et al., 2018; Stecher et al., 2018), a review of the national policy landscape reveals that
few states are “turning their backs” on educator effectiveness policies as a strategy for educational reform (Doherty & Jacobs, 2015, p. ii). Indeed, while the Every Student Succeeds Act (ESSA, 2015) removed federal incentives and provided greater flexibility to states in determining whether and how to evaluate teachers, most state legislatures have opted to maintain evaluation policies passed during the RttT years (Education Commission of the States, 2018). This, coupled with the reality that researchers and policy makers are just now beginning to grapple with how to make use of data gathered during the implementation of large-scale educator evaluation policies over the past decade to inform teacher quality efforts (Doherty & Jacobs, 2015; Kraft & Gilmour, 2017; Santelices et al., 2017), presents a need for continued research regarding educator evaluation policies such as SB 10-191.

Problem Statement

Much of the impetus for passing large-scale educator evaluation policies as part of RttT era reform efforts, apart from the potential funding opportunities associated with the federal program, derived from research findings that showed minimal attention to distributions of educator quality within pre-existing evaluation systems in public schools (Kraft & Gilmour, 2017; Weisberg et al., 2009). Such research showed that in addition to relying heavily on binary evaluation ratings (i.e. “satisfactory” or “unsatisfactory”), more than 90% of teachers received the highest ratings, while less than 1% were rated unsatisfactory (Weisberg et al., 2009).

Indeed, during the years leading up to the passage of SB 10-191 in Colorado, scholars and policy makers increasingly began to define the problem of teacher quality as
a problem of practice stemming from weak processes for supervising and evaluating teachers. Specifically, they criticized conventional methods of evaluation (Marshall, 2005) while arguing that schools have been indifferent to measuring teacher quality, except when needing to remove a teacher (Weisberg et al., 2009). As an alternative, scholars and policy makers proposed new teacher evaluation systems that place greater accountability on teachers and schools to measure teacher quality in ways that directly account for student learning outcomes (Close et al., 2018).

It was within this zeitgeist of scholarly calls for reforming teacher evaluations that Colorado enacted SB 10-191 in 2010. Since that time, additional research has begun to examine the relationships between various measures of teacher quality and the extent to which such measures can be isolated from school context factors, which are not under the direct control of the teacher (Martinez, 2013; Milanowski, 2011; Santelices et al., 2017). While still early in its development, this preliminary research has yielded mixed findings around the correlations between measures of teacher quality that focus on teacher qualifications, practices, and effects on student performance (Milanowski, 2011; Santelices et al., 2017). Moreover, recent findings suggest that although standards-based measures may be more effective in isolating teacher quality assessment from contextual factors not under the direct control of the teacher, neither standards-based nor teacher effect measures are entirely context-free (Santelices et al., 2017). As policy makers and school reformers continue to debate whether large-scale teacher evaluation policies like SB 10-191 should be mended or ended, it is important to not only consider the integrity
of various teacher quality measures, but also to consider the role of school context on these measures within teacher evaluation.

**Purpose Statement**

In an effort to contribute to the empirical literature around measuring teacher quality, and to provide a basis for deepening policy discussions around educator evaluation reform, the purpose of this study is to investigate the relationships between various measures of teacher quality and the effects of school context on teacher evaluation ratings under Colorado SB 10-191. The specific measures of teacher quality used in the study were based on those most commonly observed within the literature and evaluation policy frameworks and included teacher qualifications (level of education, years of experience, and licensure type), assessment of teacher practice, and teacher effect estimates based on student assessment results, as mandated under SB 10-191. The effects of school context were also measured using the rural/non-rural classification of the school.

As suggested by Santelices et al. (2017), investigating the relationships between various measures of teacher quality can contribute to policy and practice by providing a more nuanced understanding of teacher effectiveness and indicating whether efforts to enhance teacher effectiveness can be best addressed through educational policy, initial educator preparation, and/or ongoing professional development. In addition, an analysis of the relationships between measures of teacher quality and context variables can be used to strengthen the validity of evaluation instruments by gauging the extent to which constructs of teacher quality can be measured in ways that minimize construct-irrelevant
factors such as context variables that lie outside of the direct control of the teacher (Gregory, 2007; Santelices & Taut, 2011; Santelices et al., 2017). In the case of Colorado, this type of analysis also holds promise for assessing the effects of SB 10-191 on measures of teacher quality within districts that are largely defined by their geographical contexts. Such information can be used by policy makers and practitioners as refinements and changes are made to the Colorado State Model Evaluation System (CDE, 2019a).

To conduct this analysis, data collected by CDE as part of its annual human resource and district accountability collections was analyzed. These collections are required for all Colorado public school districts and provide a comprehensive data source for which to conduct analyses on teacher quality. The quantitative analytical data set was comprised of statewide information on each of the teacher qualifications noted above as well as evaluation ratings for each of the quality standards defined by SB 10-191, which include a teacher effect estimate based on student assessment results. The data set also included information on each school’s geographic classification as defined by CDE.

As suggested by Brewer and Kuhn (2010) for research involving relationships between independent and dependent variables after an action has already occurred, this study employed a quantitative causal-comparative (ex post facto) design to examine the relationships between measures of teacher quality and assess the effects of district context on those measures during the first years of SB 10-191 implementation. A series of multilevel models (MLM) using moderated mediation analysis were used to investigate these relationships, while accounting for the clustering of teachers within unique school
contexts. The model in Figure 1 illustrates the approach to the moderated mediation analysis used in this study as it pertains to the research questions.

Overall, this study contributes to the existing literature on teacher quality and evaluation by providing empirical evidence about the extent to which various measures of teacher quality relate and the influence of school context on those relationships. The findings of this study can be applied to refining existing teacher evaluation policies while considering whether additional constructs that explicitly consider context should be considered as part of large-scale educational reforms.

Figure 1

*Moderated Mediation Analysis of Teacher Qualifications, School Context, and SB 10-191 Evaluations Ratings*

**Research Questions**

The research questions to be used in this study were designed to investigate the relationships between teacher qualifications, professional practices and effects on student outcomes as measured by Colorado’s law, SB 10-191. More specifically, the research
questions sought to analyze the extent to which teacher practices, as captured by evaluation ratings under SB 10-191, mediate the relationships between teacher qualifications and SB 10-191 teacher effect estimates, while investigating whether a school district’s size and geographical context moderates the relationships between these measures of teacher quality. As such, the questions were designed to account for the clustering of teachers within unique district contexts, while analyzing the influence of such contexts on measures of teacher quality and student outcomes under SB 10-191. The research questions used in this study are as follows:

**Research Question 1:** To what extent are teacher qualifications associated with teacher effect estimates under SB 10-191?

**Research Question 2:** To what extent do teacher practices mediate the relationship between teacher qualifications and teacher effects?

2a. To what extent are teacher qualifications associated with better performance on the teacher quality standards used to measure practices under SB 10-191?

2b. After controlling for teacher qualifications, what is the relationship between teacher practice ratings and teacher effect estimates under SB 10-191?

2c. What is the relationship between teacher qualifications and teacher effect estimates under SB 10-191 after controlling for professional practice ratings?
Research Question 3: To what extent does school context moderate the relationships between teacher qualifications and SB 10-191 teacher quality measures?

Overview of Methodology

The research questions were addressed using data gathered by CDE as part of its annual human resources data collection cycle. The data set included evaluation ratings for individual teachers in all Colorado school districts using the Statewide Model Evaluation System during the last year of full statewide implementation of the original policy under SB 10-191 (2016-2017). This year of data was selected as it provides the most recent information on teacher quality in Colorado under the original structure of SB 10-191. As the research questions guiding this study do not involve analyzing changes in teacher quality over time, but rather, focus on the extent to which various measures of teacher quality relate and are influenced by school context, one year of data was deemed sufficient and appropriate for this targeted analysis.

Separate sets of analyses were used for each of the research questions through ordinal logistic regression using the Multilevel Modeling (MLM) technique. MLM is a complex form of analysis that uses regression techniques to analyze the variance in outcome variables in cases when predictor variables exist at various levels. MLM has become increasingly common for research in the educator sector where clustering of teachers and students within classroom, school, and district contexts occurs, thus contributing to variance within as well as between clusters that must be accounted for in the analysis (Osborne, 2000; Raudenbush & Bryk, 2002; Woltman et al., 2012). The
justification for using MLM in this study comes from the nature of the research questions as well as the nature of teacher evaluations, which occur within district contexts. By accounting for variance that occurs at both the teacher- and district-level, MLM provides a means of gaining more reliable estimates of the impact of individual teacher-level and school district-level variables on teacher quality.

**Overview of Teacher Evaluation in Colorado**

Prior to the passage of SB 10-191, teacher evaluation - in terms of both policy and practice - was largely considered a *local control* issue that lied within the realm of powers granted to local school boards within the state’s system of public education (State Council for Educator Effectiveness, 2011). Within the United States public education system, local control has been defined as an approach to school governance in which decisions about the education of children are made by those closest to the school itself (Uerling & O’Reilly, 1989). Colorado is one of only six states in the U.S. (Baker et al., 2013) that not only follows local control as a political tradition, but has it enshrined in the state constitution where local school boards are granted significant powers over instructional matters (Colo. Const. Art. IX, § 15) alongside the more general supervision powers granted to the publicly elected State Board of Education (Colo. Const. Art. IX, § 1). This unique political context has not only been important in the initial design of SB 10-191 (State Council for Educator Effectiveness, 2011), but continues to shape debates around implementation and changes to the policy, as well as the role of state education agencies in managing issues of teacher quality and evaluation (Meltzer, 2019; Meltzer & Fish, 2019; Schoales, 2017).
SB 10-191 Design and Goals

SB 10-191 was designed as an educational reform policy aimed at improving teacher quality and student outcomes through a State Model Evaluation System (State Council for Educator Effectiveness, 2011). The law lays out several goals for the use of educator evaluations in Colorado, including: to serve as a basis for the improvement of instruction; to enhance the implementation of programs of curriculum; to serve as a measurement of professional growth and development of licensed personnel; to evaluate the level of performance based on the effectiveness of licensed personnel; and to provide a basis for making decisions in the areas of hiring, compensation, promotion, assignment, professional development, earning and retaining non-probationary status, dismissal, and non-renewal of contract (Colorado Senate, 2010, p. 2).

In pursuit of these goals, the State Board of Education was directed to promulgate rules for educator evaluations based on the recommendations by a council appointed by the governor. These rules included a common definition of teacher effectiveness, as well as standards-based criteria (professional practices) for evaluating teachers using the State Model Evaluation Teacher Rubric as developed by CDE (CDE, 2015). The law also mandated that at least 50% of each educator’s evaluation shall be based on the academic growth of students using a combination of state and local assessment data as measures of student learning (Colorado Senate, 2010). Figure 2 displays the composition of a teacher’s overall evaluation rating based on SB 10-191 requirements.
Based on this distribution and using a scoring system developed by CDE, teachers are assigned an overall \textit{effectiveness rating} of ineffective, partially effective, effective, or highly effective. This rating provides the basis for making personnel decisions under the law, including earning and retaining non-probationary status (Colorado State Board of Education, 2019).

\textbf{Teacher Quality Standards}

The original rules adopted by the State Board of Education following the passage of SB 10-191 included six quality standards on which teacher evaluation ratings were to be based and were written as follows:

- Quality standard I: Teachers demonstrate mastery of and pedagogical expertise in the content they teach. The elementary teacher is an expert in literacy and mathematics and is knowledgeable in all other content that he or she teaches (e.g., science, social studies, arts, physical education, or world languages). The
secondary teacher has knowledge of literacy and mathematics and is an expert in his or her content endorsement area(s).

- Quality standard II: Teachers establish a safe, inclusive and respectful environment for a diverse population of students.
- Quality standard III: Teachers plan and deliver effective instruction and create an environment that facilitates learning for their students.
- Quality standard IV: Teachers reflect on their practice.
- Quality standard V: Teachers demonstrate leadership.
- Quality standard VI: Teachers take responsibility for student academic growth (CDE, 2015).

Quality standards I through V were further broken down into elements and professional practices within the State Model Evaluation Rubric. In total, the original rubric was comprised of 27 elements and 309 professional practices (CDE, 2018b).

During the 2016-2017 school year, CDE initiated a revision process of the State Model Evaluation System for teachers that included a reworking of the original quality standards and elements based on implementation data and feedback from the field of educators and administrators in Colorado’s public schools (CDE, 2019b). This process resulted in the refinement of the State Model Evaluation rubric and an overall reduction of the six quality standards noted above to five quality standards (Quality Standards IV and V were combined in to a new “Professionalism” Standard). The revision also included a reduction of elements and professional practices within the state model evaluation rubric, which, following approval by the State Board of Education in 2018, is
currently is comprised of 17 elements and 165 professional practices (CDE, 2018b). As insufficient data exists to accurately analyze the revised structure, this study draws only from data collected during the original implementation of the State Model Evaluation System under SB 10-191.

**Measures of Student Learning**

In order to meet the SB 10-191 requirement that 50% of each educator’s evaluation be based on student academic growth, (quality standard VI in the original State Board rules and quality standard V in the revised rules), school districts must annually compile measures of student learning (MSLs) that include the results of state and local assessments. The State Board rules further specify that such measures must include:

- A measure of individually attributed Student Academic Growth, meaning that outcomes on that measure are attributed to an individual licensed person.
- A measure of collectively-attributed Student Academic Growth, whether on a school-side basis or across grades or subjects, meaning that outcomes on that measure are attributed to at least two licensed personnel (e.g., measures included in the school performance framework, required pursuant to section 22-11-204, C. R. S.).
- When available, statewide summative assessment results; and
- For subjects with annual statewide summative assessment results available in two consecutive grades, results from the Colorado Growth Model (Colorado State Board of Education, 2019, p. 18).
As a stand-alone, indirect measure of student learning, MSL ratings (i.e. Standard VI or V) provide an estimate of each teacher’s contributions to student outcomes under the Colorado State Model Evaluation System. Using a scoring formula developed by CDE that is based upon the extent to which individual teachers meet annual growth targets, the combined MSLs yield a rating that, when coupled with the professional practice ratings, make up the overall evaluation score of a teacher for a given year.

**Distribution of Teacher Quality Under SB 10-191**

Few empirical studies have been conducted on SB 10-191 and the research that has been conducted has focused primarily on analyzing the overall distribution of educator evaluation ratings across Colorado. Such analysis conducted during the initial pilot of SB 10-191 revealed that the system produced some distributions of ratings across categories of teacher performance within the standards and elements captured in the teacher evaluation rubric (CDE, 2013a). Specifically, teachers were found to score highest on Standards 2 (Establish Environment) and 5 (Demonstrate Leadership) and lowest on Standard 3 (Facilitate Learning). These findings suggest that the system is able to produce data that can be used by school districts to inform instruction, implementation of programs, and personnel decisions, which was one of the central goals of SB 10-191. Yet, the initial report also revealed considerable variance between individual pilot school districts with some having 100% of teachers rated as proficient or higher while others had just over half of teachers rated proficient or higher (CDE, 2013a). Data from later pilot reports has yielded similar results with 98% of pilot district teachers being rated as proficient or higher during the 2014-2015 year (CDE, 2016). While acknowledging that
this data may in fact reflect the true performance levels of teachers within the pilot school districts, CDE has also questioned whether these results were a function of low comfort levels of individual evaluators with assigning low ratings or having low levels of familiarity with the teacher evaluation rubric (CDE, 2016).

**Conceptual and Theoretical Frameworks**

The conceptual framework used in the design of this study draws from the field of educational effectiveness research (EER), which is concerned with addressing questions about the ways in which teachers and schools can contribute to educational outcomes, while reducing the variations in outcomes that have been attributed to students’ background characteristics (Kyriakides et al., 2018). A key component of EER as it is developed over the past three decades has been the use of multilevel quantitative models to investigate the relationships among variables occurring at different levels (student, class, school), while also examining the role of contextual factors on measures of effectiveness (Cervini, 2009; Santelices et al., 2017). Specifically, the *dynamic theory of educational effectiveness* developed by Creemers and Kyriakides (2006, 2008, 2018) provides a framework to guide research and has been operationalized in this study within the development of the research questions, the selection of methodology, and was used to support interpretation and analysis of the findings. With regard to the selection of variables to be analyzed, this study also drew upon several frameworks in the literature that focus on measures of teacher quality and rurality as a relevant context-defining factor in research on educational outcomes.
Dynamic Theory of Educational Effectiveness

In general, the dynamic theory of educational effectiveness examines the role of teachers and schools on educational outcomes while taking into account the complexity of educational systems in which clustering of students and teachers occurs, and utilizing quantitative approaches that allow researchers to examine the influence of variables within and across levels of an educational system (Kyriakides et al., 2018; Nilsen & Gustafsson, 2016). The theory is concerned not only with addressing questions about the effect of teachers and schools on student achievement - referred to as the quality dimension of effectiveness, but also with investigating the extent to which educators can reduce variations in student outcomes, particularly those attributed to student background characteristics - referred to as the equity dimension of effectiveness (Kyriakides et al., 2018).

In developing the theory, Creemers and Kyriakides (2006) note that an absence of theoretical models, combined with narrow definitions of effectiveness based on student performance on standardized assessments (typically in math) have historically led to incomplete “education production” models that oversimplify the relationships between inputs and outputs in education. For example, while reducing student-to-teacher ratios or increasing education funding has been shown to be positively associated with increases in student outcomes, neither of these interventions necessarily result in higher student outcomes, suggesting that the relationship between variables is more complex than once assumed. To address these issues, the dynamic theory of educational effectiveness posits that models of educational effectiveness must: a) be multilevel in nature, b) be based on
the assumption that the relation of some effectiveness factors with achievement may be curvilinear, c) illustrate the dimensions upon which the measurement of each effectiveness factor should be based, and d) define relations among the effectiveness factors (Creemers & Kyriakides, 2006).

**Teacher Quality Frameworks**

In addition to the dynamic theory of educational effectiveness, which provides a framework for conceptualizing relationships between effectiveness factors at various levels of educational systems, this study also drew upon several frameworks for understanding key components of teacher quality and their relations to student outcomes. This included (a) research on the observable teacher characteristics that may impact student performance such as education level, license type, and years of experience (Goe, 2007; Goe et al., 2008; Hanushek & Rivkin, 2006; Rice, 2003); (b) research on teacher quality as defined by professional practices based on professional (i.e. quality) standards and measured through evaluation processes (Danielson, 2007; Darling-Hammond et al., 2011; Marshall, 2005); and (c) value-added estimates based on student performance on standardized assessments (Bill & Melinda Gates Foundation, 2013; Millman 1997). In the context of this study, teacher quality is defined by a combination of measures reflecting a teacher’s training and experience (*inputs*), instructional practices (*processes*), and direct effects on student learning (*outputs*) (Goe, 2007; Nielson & Gustafsson, 2016).

A distinctive feature of SB 10-191 (and other teacher evaluation laws passed during the RttT era) is that it includes standards that measure the processes and outputs domains of teacher quality, but does not include any direct measures around the inputs.
associated with teacher quality. For example, neither a teachers’ education level nor years of experiences are considered in a teacher’s overall evaluation rating. Such an omission of a key construct of teacher quality within a framework for teacher evaluation not only reflects a potential dissonance between theory and policy, but also provides further justification for investigating the extent to which teacher qualifications and teacher evaluation ratings are related. Such an investigation can inform policy development by indicating whether input factors should be considered along with process and output factors in teacher evaluation, particularly when measured within dynamic educational systems that are situated within unique school contexts.

**Conceptualizing School Context**

In discussing the challenges associated with conducting educational research to inform policy decisions, Berliner (2002) notes that the “power of contexts” and the “ubiquity of interactions” that emerge from the complex activities occurring within schools make educational research the “hardest science of all” (p. 18). Adding to these challenges is the task placed before educational researchers, of attempting to define and conceptualize educational contexts in what Labaree (2003) has termed, “marshy epistemological terrain” (p. 14). Examples of “evidence-based practices” which have been found to be context-dependent in terms of measuring their “effectiveness” through “scientific research” include instructional and curricular models, assessments, and behavioral interventions (Berliner, 2002).

Despite these challenges, several scholars have drawn upon existing theories such as ecological theory and systems theory to develop constructs of educational contexts that
can be applied to educational research (Frisby & Reynolds, 2005; Stelmach, 2011). Like the dynamic theory of educational effectiveness, these theories conceive of education as a complex process that is shaped by economic, social, and political forces that surround schools and operate at macro-, mezo-, and micro- levels (Stelmach, 2011). Within this realm of scholarship, rural education studies, which focus on the impact of geographic isolation on education, have emerged as increasingly relevant in studying the effects of such issues as access to resources, teacher recruitment and retention, and poverty on efforts to close equity gaps through various educational policies and reforms (Arnold et al., 2007; Cloke, 2006; Coladarci, 2007).

In Colorado 148 (83%) of the state’s 178 school districts are classified as rural (at least 25 miles from urban area with 6500 students or less) or small rural (at least 25 miles from urban area with 1000 students or less) by CDE (CDE, 2018b). Thus, a school district’s rural status, as defined by its size and geographic location, has been established as an important factor to consider in this study as it can provide a basis for operationalizing school contexts within the state of Colorado while providing insights about the ways in which those contexts relate to measures of teacher quality and effectiveness. In addition to making theoretical contributions to EER and research on teacher quality, the results and conclusions from this study also hold promise for informing policy and practice discussions related to teacher evaluation under SB 10-191.

Application of Theoretical Models and Hypotheses

Figure 3 illustrates the application of the dynamic theory of educational effectiveness, teacher quality frameworks, and research on school contexts to the
conceptual design of this study.

**Figure 3**

*Conceptual Framework developed for this study using a dynamic model of educational effectiveness.*

Based on this model, teacher qualifications (inputs) such as education level, license type, and years of experience are hypothesized to be related to performance on SB 10-191 teacher quality measures including professional practices (processes) and measures of student learning/teacher effect estimates (outputs). By reflecting instructional practices that occur on a consistent basis between teachers and students, SB 10-191 professional practice measures (processes) are hypothesized to mediate the relationship
between teacher qualifications (inputs) and SB 10-191 measures of student learning/teacher effect estimates (outputs). Finally, school district context, as reflected by rurality, is hypothesized to moderate the relationship between teacher qualifications and measures of teacher quality under SB 10-191. In other words, the strength of the relationships between teacher qualifications and measures of teacher quality under SB 10-191 is expected to be different between groups of teachers based on the school district’s rurality status. Thus, in line with similar research examining the role of educational contexts on teacher quality measures (Santelices et al., 2017), neither of the measures used to evaluate teachers under SB 10-191 are hypothesized to be context-free.

Rationale and Significance

The rationale for this study stems from both theoretical and policy-driven needs to better understand constructs of teacher quality, how they are measured, and the influence of school district contexts in evaluating teacher performance. From a theoretical standpoint, this study aims to contribute to the development of dynamic models of educational effectiveness that can be used to measure and understand relationships between school- and teacher-level factors on student outcomes (Creemers & Kyriakides, 2006). Such inquiry holds promise not only for addressing issues of quality and equity in education, but also for informing policies and practices related to teacher evaluation and human resource management within schools.

From a policy perspective, this study holds significance for providing insight around the constructs used to define teacher quality and evaluate teacher performance under SB 10-19. Indeed, as the policy continues to be refined, such insights can be useful
in determining the extent to which other factors (i.e. teacher qualifications, context variables) should be considered within evaluation policy frameworks. Finally, this study will also address important questions about whether the measures used to evaluate teacher performance under SB 10-191 are context-free or whether additional consideration should be given to the role of school district geographical location in the design of teacher evaluation policies (Santelices et al., 2017). These findings can also be useful in the design of other types of large-scale reform policies that seek to address issues of quality and effectiveness in education.

**Organization of the Dissertation**

Chapter II provides a review of the extant literature concerning teacher quality from both theoretical and policy-based perspectives. Attention is given to the role that recent scholarship and the politics of accountability have played in the design of large-scale teacher evaluation policies and SB 10-191 in particular. The conceptual frameworks used in the design of this study are also reviewed in terms of the theoretical advances made, thus far, and areas in need of additional research.

Chapter III explains how MLM will be applied to investigating the research questions concerning the relationships between measures of teacher quality and school district contexts in Colorado under SB 10-191. Descriptions of the variables used in this study are provided, along with model specifications detailing how ordinal logistic regression analysis was applied to analyzing the variables. Limitations and delimitations are discussed in terms of how they affect the validity of the study and interpretation of the results.
Chapter IV provides descriptive statistics on the research sample used in this study and presents the results of the ordinal logistic regression for each phase of the moderated mediation analysis.

Chapter V employs the findings of the moderated mediation analysis to address each of the research questions and consider the implications of the findings on educational policy and practices related to teacher evaluation.
CHAPTER II
LITERATURE REVIEW

The extant literature addresses issues of teacher quality through multiple lenses and includes research spanning several decades on the influence of initial qualifications (i.e. preparation, licensure), instructional practices, and ongoing training and experience on student learning outcomes (Goe, 2007; Goe et al., 2008; Nilsen & Gustafsson, 2016). But despite nearly universal agreement that teacher quality matters, scholars have yet to achieve a clear consensus around which aspects of teacher quality matter most, or how best to measure and evaluate teacher performance (Goe et al., 2008).

Adding to the challenges of defining and measuring teacher quality are the high-stakes that have become attached to teacher performance and evaluation through policies accompanying large-scale education reforms like the Bush administration’s No Child Left Behind Act of 2001 (NCLB, 2002) and the Obama administration’s RttT (RttT, 2011) Initiative. These policies have sought to address issues of teacher quality through standards and accountability measures that place high expectations and on teachers and schools in ways that, some have argued, frustrate practices in evaluating teacher quality and performance (Ramirez et al., 2014).

Based on this intersection of research and policy that has characterized modern discourse on teacher quality and evaluation and contributed to the development of policies like SB 101-91, this literature review draws upon studies completed since the dawn of the standards and accountability era under No Child Left Behind in 2001 (NCLB, 2002). It applies the framework developed by Goe (2007) which defines teacher
quality in terms of inputs, processes, and outputs, and provides a synthesis of major findings in the literature while highlighting the limitations, gaps, and deficiencies that have informed the design of this study and the research questions. It also discusses the developing body of research that examines the role of school district context on measures of teacher quality under modern large-scale evaluation systems.

**Historical Context: Issues of Teacher Quality in Education Reform**

Over the past quarter century, issues of teacher quality have played an expanding role in shaping education policies in the United States and have been key drivers of educator evaluation policies like SB 10-191 (Anagnostopoulos, Lingard, & Sellar, 2016). Indeed, since the beginning of the standards and accountability era spurred by the publication of A Nation at Risk (United States, 1983), the idea that teacher quality could be enhanced through increased accountability became central to the design and promotion of federal and state education policies (U.S. Department of Education, 2008). Perhaps nowhere was this focus more prominent than in the passage of the federal No Child Left Behind Act of 2001 (NCLB, 2002), which mandated that states adopt new academic standards and assessments in order to close achievement gaps between various subgroups of students while ensuring universal proficiency in reading and math by the year 2014 (Klein, 2015). In addition to these mandates, a key provision of NCLB specified that all school districts must take steps to ensure that teachers become “highly qualified” by 2005-2006 school year (NCLB, 2002). The criteria established by the law for designating teachers as “highly qualified” included: a) a requirement that teachers hold a minimum of a bachelor’s degree, b) a requirement that teachers meet state
qualification and licensing criteria for the grade levels and subject areas in which they provide instruction, and c) a requirement that teachers demonstrate knowledge of the subjects they teach (NCLB, 2002).

Despite efforts to enhance teacher quality and student performance through the provisions of NCLB, many researchers deemed it a failed federal policy, not only because it fell short of meeting its goals of universal proficiency in reading and math, but also because it produced unintended consequences such as incentivizing “teaching to the test,” and the artificial inflation of test results by school districts and state agencies (Close et al., 2018). Nonetheless, issues of teacher quality and effectiveness lingered as key topics of debate within subsequent educational reform efforts (Anagnostopoulos et al., 2016) and became central to many states’ strategies for accessing federal funding under subsequent education reform efforts. Under President Obama’s (RttT, 2011) initiative, states could compete for federal grants by developing assessments of teacher quality (i.e. teacher evaluations) that included measures of teacher performance on professional standards as well as measures of student achievement through what became known as “value-added” models of evaluation (Close et al., 2018; Goe, 2007, Goe et al., 2008; RttT, 2011). These value-added models (VAMs) employed advanced statistical methods in attempts to track the contributions of individual teachers to student achievement over time, while attempting to control for prior learning and student background characteristics (Amrein-Beardsley, 2008).

By 2012, when it became clear that the 100% proficiency levels established by NCLB would not be met, states were also provided with opportunities to obtain waivers
from the law by adopting reforms, some of which included educator evaluation systems that incorporated measures of student achievement through VAMs or other methods used to directly link measures of teacher quality to student performance (Close et al., 2018). The following section provides a brief overview of the literature on evaluation of teacher quality that has contributed to policies like SB 10-191 and highlights some of the gaps that have informed the design of this study.

**Measuring Teacher Quality through Evaluation**

While the conventional wisdom behind teacher evaluation has been that it contributes to “better” teaching, which in turn can lead to “better” learning (Bill & Melinda Gates Foundation, 2013), there has been considerable disagreement amongst scholars and policy makers about the best ways to evaluate teachers. During the years leading up to the passage of large-scale teacher evaluation policies like SB 10-191, scholars and policy makers began to increasingly define the problem of teacher quality as a problem of practice stemming from weak processes for supervising and evaluating teachers (Marshall, 2005). They also framed weak teacher evaluations as a problem of equity that contributed to low performance of students, particularly those in high-poverty schools (Weisberg et al., 2009).

Research conducted prior to the RttT Era led some scholars to argue that conventional methods of teacher evaluation, which relied on micro-evaluations of teacher practice conducted by principals through formal observations of one or two lessons, were drastically flawed and gave little focus to student learning (Marshall, 2005). Such systems were characterized as little more than a “pro forma” processes that had very little
influence improving teacher quality (Marshall, 2005, p. 729). As an alternative, scholars argued that evaluation systems should consist of frequent classroom observations by a principal, accompanied by face-to-face feedback, and should encourage teacher collaboration on common units and assessments for measuring student learning (Darling-Hammond et al., 2011; Marshall, 2005). They also recommended the use of rubrics over conventional evaluation instruments (i.e. checklists) as they allow for clear feedback with respect to standards at varying levels of proficiency (Darling-Hammond et. al, 2011; Marshall, 2005).

Other scholars went beyond a limited focus on weak practices in their calls for reform, charging that schools have historically been “indifferent to instructional effectiveness – except when it comes time to remove a teacher” (Weisberg et al., 2009, p. 2). These scholars coined the term *widget effect* to describe the “tendency of school districts to assume classroom effectiveness is the same from teacher to teacher…except teachers whose performance is so egregiously poor as to warrant dismissal” (p. 2). As a remedy to such issues, they argued in favor of stronger state policy frameworks that considered a teacher’s effectiveness in key decisions such as recruitment, hiring, tenure, school placement, and dismissal. They also recommended that such frameworks include clear performance standards with multiple rating options, training and accountability systems for principals, and integration of evaluation systems with human resource functions such as teacher assignment, professional development, and dismissal (Weisberg et al., 2009).
Perhaps the most compelling research on teacher evaluation that combined both elements of practice and effectiveness came from the Measures of Effective Teaching (MET) Project, launched by the Bill and Melinda Gates Foundation in 2009. This 3-year study was designed to determine whether it is possible to identify “great teaching” by combining three types of measures: classroom observations, student surveys, and student achievement gains (Bill & Melinda Gates Foundation, 2013). The study’s findings suggested that reliable measures could be achieved by combining multiple measures, albeit only by attending to a careful balance between measures and allowing multiple observers to evaluate teacher practice and provide regular feedback (Bill & Melinda Gates Foundation, 2013).

In addition to these scholarly calls for reforming teacher evaluation practices, state-level policy advocates of SB 10-191 also framed the issue of teacher quality as a matter of equity and one that could be addressed through increased accountability by requiring schools to base 50% of each teacher’s evaluation on student assessment results. This enhanced focus on accountability also led to a provision of the bill allowing teachers to lose tenure status if not consistently contributing to student learning as measured by gains on such assessments (Colorado Senate, 2010).

Despite widespread agreement that change was necessary to improve teacher evaluation, not all agreed upon the role of student assessments in measuring teacher quality. Indeed, some scholars criticized teacher evaluation systems that relied on single assessments as misguided, while arguing that even when multiple measures are used, the task of measuring the impact of individual teachers on student learning is inherently
complex and controversial (Hinchey, 2010). Others honed criticism on emergent value-added models (VAMs), which examined changes in student test scores over time, and argued that such models were highly unstable and incapable of yielding valid information on the influences affecting student progress (Darling-Hammond et. al, 2011). Such sentiments were echoed by opponents of SB 10-191, who argued that, while “everyone agrees that teachers should be held accountable for their performance, even teachers...virtually no one has any idea how to measure that performance” (Scott, 2010, para. 9).

It is within this context of ongoing education reform that Colorado enacted SB 10-191 in 2010. It is also within this context that debates began to emerge within educational research and policy circles around how to best define teacher quality and the value of teacher evaluation systems as mechanisms for measuring teacher performance and improving student outcomes (Barnette & Amrein-Beardsley, 2011; Engdahl, 2010; Goe et al., 2008; Scott, 2010).

**Teacher Quality**

While NCLB emphasized teacher characteristics associated with initial training and licensure in defining teacher quality, subsequent education reform policies have placed greater emphasis on the role of instructional practices and individual teacher effects on student learning as measured by VAMs. These shifts have been accompanied by research that has attempted to clarify definitions of teacher quality while exploring the relationships between various facets of teacher quality and student outcomes. Drawing upon this research Goe (2007) developed a framework for teacher quality that continues
to guide researchers and policy makers around the globe (Nilsen & Gustafsson, 2016; Santelices et al., 2017). According to the framework, teacher quality is comprised of teacher qualifications and other characteristics (inputs) that influence teacher practices (processes) which in turn determine a teacher’s effectiveness in contributing to student achievement gains (outcomes). Figure 4 provides a visualization of Goe’s (2007) framework while illustrating how the various facets of teacher quality are related.

**Figure 4**

*Framework for Teacher Quality adapted from Goe (2007)*

In developing the framework, Goe (2007) notes the importance of distinguishing between *teacher* quality and *teaching* quality, which are often conflated in educational research and policy and have led to the assumption that *teacher* quality automatically
ensures teaching quality. She also notes that distinguishing between what teachers do in classrooms and what students learn in classrooms is important for defining successful teaching and gauging teacher effectiveness (Goe 2007). What follows is a more in-depth analysis of the scholarship that has contributed to each of the facets of teacher quality as defined within Goe’s (2007) framework for teacher quality.

**Teacher Qualifications and Characteristics**

Within the framework for teacher quality developed by Goe (2007), teacher qualifications and characteristics are considered inputs that teachers bring with them to the classroom and, while theorized to influence other facets of teacher quality, they exist independently of instructional practices or student outcomes. Specifically, teacher qualifications refer to credentials (which include certifications and measures of subject matter knowledge), education levels, teacher preparation and licensure pathways, education levels (degrees and professional development), and prior experience that teachers bring to their work (Goe, 2007; Santelices et al., 2017). Teacher characteristics are defined as other assigned attributes that characterize the teachers themselves, such as race or gender. While research on teacher qualifications has been abundant since the advent of NCLB, research on teacher characteristics has been scarce due to ethical and practical considerations that lie outside a teachers’ or researchers’ ability to change (Goe, 2007). Thus, the focus for this literature review is on research and variables that fall within the teacher qualifications category of inputs.
Credentials

As previously noted, NCLB (2002) placed heavy emphasis on teacher credentials in defining the criteria for the designation of “highly qualified.” Despite the weight attached to credentials under the law, however, research has yielded mixed results with regard to how they relate to student achievement (Hanushek & Rivkin, 2006; Kennedy, 2008; Rice, 2003; Santelices et al., 2017). Indeed, some studies examining the role of various credentials such as National Board Certification and subject-matter knowledge on student achievement have found that correlations vary across grades and subject areas (Baumert et al., 2010; Betts et al., 2003; Blömeke & Delaney, 2014; Carr, 2006; Cavalluzzo, 2004; Clotfelter et al., 2006; Darling-Hammond, 2000; Darling–Hammond et al., 2005; Goldhaber & Anthony, 2005; Hill et al., 2005). Not all scholars have agreed however, as some studies have found no significant links between credentials and student gains on standardized test scores (Gordon et al., 2006; McColsky et al., 2005; Rivkin et al., 2005).

The mixed findings related to credentials and subject-matter knowledge have contributed to criticisms of credentials-only approaches to measuring teacher quality and several scholars have cautioned against reliance on such “paper qualifications” arguing that holding a particular credential is not a guarantee that teachers will be effective in their classrooms (Barnett & Amrein-Beardsly, 2011; Goe, 2007). Likewise, the mixed findings of research on credentials have contributed to demands for research that accounts for additional variables and can explain greater proportions of student performance relative to teacher qualifications (Kennedy, 2008; Santelices et al., 2017).
Teacher Education Levels

Studies examining the influence of teachers’ education levels on student achievement have found that holding a higher degree such as a masters or doctorate can contribute to higher student achievement, though results have been marginal and have varied depending on subject matter (Betts et al., 2003; Clotfelter et al., 2006; Clotfelter et al., 2010; Obonyo et al., 2018). More promising effects have been observed in studies focusing on professional development rather than degrees, where training that enhances teachers’ content knowledge and pedagogical content knowledge has been found to be a significant predictor of student achievement in math and science (Harris & Sass, 2007; Newsome et al., 2019; Rice, 2003).

In spite of the research showing weak correlations between teacher education levels (degrees) and gains in student achievement, there continues to be a reliance on such factors as determinants of hiring, salaries, eligibility for certifications, and other types of human resources decisions made by states and school districts (Goldhaber, 2002). As with credentials, this has prompted calls for more refined research that considers how education levels work in conjunction with other measures of teacher quality to influence student learning (Hanushek & Rivkin, 2006).

Preparation and Licensure Pathways

An emerging body of research within the realm of teacher qualifications has focused on investigating the extent to which various pathways to teacher preparation and certification matter to student achievement (Jang & Horn, 2017). These studies, too, have yielded mixed results with some finding that traditional teacher education and
certification contributes more to student achievement (Clotfelter et al., 2010; Darling-Hammond et al., 2002; Zientek, 2007), while other studies have found that alternative teacher preparation and certification programs are equally or even more effective in contributing to student achievement (Hanushek et al., 2005; Henry et al., 2014; Kane et al., 2006; Lincove et al., 2015). As with other research on teacher qualifications, some studies have found that the extent to which teacher preparation and licensure pathways matter depends on the subject area being studied and while alternatively trained teachers may start out behind their traditionally-prepared colleagues, they are able to catch up to them by their second year on the job (Boyd et al., 2005).

Given the proliferation of alternative pathways to teaching that has occurred over the past decade, along with mixed findings from studies on preparation and licensure pathways, there remains a need for ongoing research that examines the influence of teacher preparation and certification pathways on aspects of teacher practice that affect student achievement (Fry & Anderson, 2011; Jang & Horn, 2017; Woods, 2016). There also remains a need to understand the extent to which distributions of traditionally and alternatively trained teachers varies across different types of schools and how those distributions, in turn, affect student achievement within diverse school contexts (Fry & Anderson, 2011; Shuls & Ritter, 2013).

**Experience**

Another strand of research to emerge around teacher qualifications has focused on the relationship between teacher experience and student achievement. Here, too, individual studies have produced mixed results with some studies suggesting that
experience, particularly within the first five years of teachers’ careers, can have positive effects on their ability to teach (Betts, Zau, & Rice, 2003). Other studies, however, have found no significant contributions of experience to student achievement (Carr, 2006; Gallagher, 2004). Despite these disagreements among individual studies, several meta-analyses on studies conducted on teacher qualifications have agreed that years of experience is an important factor in assessing teacher quality and should play a role in shaping educational policies around teacher quality (Hanushek & Rivkin, 2006; Rice, 2003; Santelices, 2017).

Teacher Practices

Goe (2007) defines teacher practices as what teachers do in their work both in and out of the classroom. This includes planning, instructional delivery, classroom management, interactions with students, and other processes associated with teaching and learning. Prior to the onset of large-scale teacher evaluation policies like SB 10-191, much of the research on teacher practices has relied on data collected through local observation protocols and evaluation instruments leading some scholars to question the integrity of data and methods used within studies of teacher practice (Goe, 2007). With the development of widely-adopted instructional and evaluation frameworks that followed implementation of large-scale evaluation policies, however, research on teacher practices has been deemed to be more valid and comparable across state and local contexts (Milinowski, 2011). Still, questions remain around the extent to which contextual factors such as poverty, infrastructure, and resources influence measures of
teaching practices through standards-based measures and protocols (Bryk et al., 2012; Graue et al., 2012; Santelices, 2017).

Much of the research on teacher practices has drawn from data collected through the widely adopted Framework for Teaching developed by Danielson (1996, 2007), which measures teacher practices based on four domains: planning and preparation, the classroom environment, instruction, and professional responsibilities. The findings have shown positive correlations to exist between teacher evaluation scores on such factors as planning and preparation, classroom management, instruction, and student achievement (Borman & Kimball, 2005; Heneman et al., 2006; Holtzapple, 2003; Milanowski, 2004). In some cases, the research has also suggested that teacher evaluation scores on the domains measured by evaluation frameworks are stronger predictors of student achievement than are teacher education and experience (Kimball et al., 2004).

Despite growing consensus that teacher practices can serve as a valid measure of teacher quality, the advent of large-scale teacher evaluation policies like SB 10-191 has prompted calls for additional research that examines convergent validity between measures of teacher practice with other measures of teacher quality such as VAMs used for evaluation purposes (Santelices & Taut, 2011; Santelices et al., 2017). Gregory (2007) defines convergent validity as the extent to which two tests of a similar construct are correlated and suggests that moderate to high correlations ($r \geq 0.5$) can serve as evidence of convergent validity. Some of the studies that have measured convergent validity between measures of teacher practice and VAMs have shown small to moderate correlations ($0.1 < r < 0.5$) between the two constructs (Hill et al., 2010; Jacob &
Lefgren, 2005; Milanowski, 2011; Pianta et al., 2008; Santelices, 2017), while others have shown higher correlations between 0.55 and 0.7 (Schacter & Thum, 2004). These disparate findings suggest that convergent validity may be dependent on the specific design of large-scale teacher evaluations policies or the ways in which they are implemented, thus justifying the need for additional research that considers how contextual factors influence the relationships between constructs of teacher quality.

**Teacher Effectiveness**

A third essential component of Goe’s (2007) framework for teacher quality is teacher effectiveness, which is defined as a teacher’s contribution to student learning, most commonly measured by VAMs or other models that link data between teachers and students to track changes in student test scores over time. By relying on complex statistical models that account for students’ prior achievement and variables outside of teachers’ control such as student background and sociodemographic characteristics, VAMs have been theoretically proposed to provide fairer estimates of *teacher effects* than models that rely on simple score averages or differences (Harris & McCaffrey, 2010; McCaffrey et al., 2004). In practice, VAMs have been criticized for being unstable due to small sample sizes (Darling-Hammond et al., 2011, 2012; Hanushek & Rivkin, 2006), for having issues with test content and measurement error (Cohen, 2010; Hill, 2009), for failing to adequately account for classroom and school-level variables that lie outside of teachers’ control (Aaronson et al., 2003; Amrein-Beardsley & Holloway, 2019), and for assuming random assignment of children to schools and classrooms, which is rarely observed in reality (Rothstein, 2010).
Despite the criticisms that have been levied against VAMs in the educational research literature, many states and policy makers continue to promote their use in evaluations of teacher quality (Amrein-Beardsley & Holloway, 2019). Moreover, research continues to show moderate correlations between VAMs and other measures of teacher quality, which has led to suggestions that, rather than being rejected as tools or gauging teacher effectiveness, VAMs be refined through the use of multilevel modeling and other statistical methods that account for student, class, and school characteristics that lie outside the direct control of teachers (Everson et al., 2013; Santelices et al., 2017). It is within this realm of modern policy and scholarly debate that additional research around the role of context on measures of teacher quality has become a focus of much of the recent scholarship on teacher quality. The following section provides a review of such scholarship while highlighting areas identified for ongoing research.

**The Role of Context**

Berliner (2002) argues that the “power of contexts” is an essential component of educational research and that the “ubiquity of interactions” that occur in classrooms and schools make educational research the “hardest science of all” (p. 18). In a similar fashion, Labaree (2003) suggests that the “messy reality” of schools often contributes to the “marshy epistemological terrain” in which educational researchers must conduct their work (p. 14). Nonetheless, methodological advances in the application of Multilevel Models (MLM) have enabled researchers to operationalize school context as a variable in educational research by analyzing data that is “clustered” within classrooms, schools, and district levels (Osborne, 2000; Raudenbush & Bryk, 2002; Santelices et al., 2017;
Woltman et al., 2012). Within the literature on teacher quality, school contexts have largely been defined by variables that reflect the conditions under which teaching and learning take place. These variables have included information on organizational factors such as school leadership and human resource practices (Heck, 2006; Heneman & Milanowski, 2004), policy factors linked to teacher evaluation, supervision and tenure (Hanushek & Rivkin, 2006; Ramirez et al., 2010, 2014), socio-economic factors such as school district and student poverty levels, (Torres, 2018) and the geographic location of schools (Arnold et al., 2007; Cloke, 2006; Coladarci, 2007).

Research has shown that school context can play an important role in analyzing teacher quality, both in terms of influencing how “highly qualified” teachers are distributed (Adamson & Darling-Hammond, 2011; O’Day & Smith, 2016) and in measuring teacher quality through standards-based evaluations and VAMs (Santelices et al., 2017; Torres, 2018). Prior to the arrival of large-scaled teacher evaluation systems like SB 10-191, much of the data used to study teacher quality within varying school contexts drew from panel data compiled by states in fulfillment of NCLB requirements (Hanushek & Rivkin, 2006). These data sets primarily included information on “paper qualifications” of teachers (i.e. education levels certifications, and years of experience), while omitting information about teacher practices (Barnett & Amrein-Beardsly, 2011; Goe, 2007). This has led to gaps in the literature that have only recently begun to be addressed through more recent research that focuses on multiple measures of teacher quality gathered through large-scale evaluation systems within various school and district contexts (Garcia, 2019; Santelices et al., 2017; Torres, 2018).
Distributions of Teacher Quality by School District Context

Given the focus of NCLB on teacher qualifications (i.e. degrees, credentials, subject matter knowledge, and experience), a great deal of attention has been given to studying distributions of “highly qualified” teachers across various school and district contexts (Adamson & Darling-Hammond, 2011; O’Day & Smith, 2016). Within this research, socio-economic factors such as school-level poverty and racial composition have consistently been identified as variables impacting distributions of teacher quality, where students of color in low-income schools have been found to be three to 10 times more likely to have unqualified teachers than students in predominantly white, higher-income schools (Adamson & Darling-Hammond, 2011).

Other research has linked significant school- and district-level context factors, such as poverty, to geographic location, where rural schools have been found to have higher rates of poverty, concentrated poverty, and intergenerational poverty when compared to urban or metropolitan areas (Mattingly et al., 2011). In addition, teachers in rural schools have been found to be less likely to hold a master’s degree (Provasnik et al., 2007), more likely to teach out of their field (Lazarus, 2003), and more likely to be novice professionals (Gagnon & Mattingly, 2014; Player 2015). Scholars have also noted the influence of geographical isolation on the ability of rural school districts to recruit and retain teachers, making high rates of teacher turnover a consistent challenge for rural schools, and particularly those with large populations of minority students and students in poverty (Gagnon, 2016; Lavalley, 2018; Showalter et al., 2017; Schaefer, Mattingly, & Johnson, 2016; Sutcher et al., 2016).
Thus, a school district’s geographic location has become an increasingly significant contextual factor in studies on teacher quality where uneven distributions of teachers have been observed between remote rural schools and those located near metropolitan areas. Studies on student achievement have also revealed differential patterns between rural and non-rural schools where, historically, lower rates of literacy, advanced coursework in secondary school, and college attendance have been observed among students attending rural schools (Clarke, 2014; Lavalley, 2018).

Taken together, the research findings on teacher quality and student achievement have led some to hypothesize that teaching is more effective in non-rural contexts with lower rates of poverty and minority students (Isenberg et al., 2013; Sass et al., 2012). However, such hypotheses have rarely been directly addressed through empirical research (Torres, 2018). Furthermore, recent studies have provided some evidence that teacher practices, as measured by large-scale evaluation policies, do not always reflect the same differential patterns observed between rural and non-rural schools around teacher qualifications. For example, Garcia (2019) found that, in Colorado, teachers in small rural schools were more likely to receive higher evaluation ratings on professional standards most closely associated with classroom practice such as pedagogical content knowledge, classroom management, and instructional delivery. In a similar fashion, Torres (2018) found variance in teacher effects, as measured through VAMs in Chile’s evaluation system, which revealed patterns of teacher quality that did not always fall along the same socio-economic divide as measures comprised of teacher qualifications alone.
These findings imply that measures of teacher quality that incorporate data beyond NCLB qualifications can be useful for providing a more nuanced understanding of how teachers are distributed between various school contexts. Nonetheless, whether any measures of teacher quality can be deemed entirely “context free” has been brought into question by recent research, which has found that school and municipality (i.e. district) context variables can explain as much as 30% of the variance in teacher quality as measured by standards-based assessments and VAMs within large-scale evaluation systems (Santelices et al., 2017). These findings have given rise to calls for additional research that further explores the role of context in modern teacher evaluation schemes, while promoting questions about whether such schemes should more explicitly incorporate context variables in gauging teacher quality through evaluation (Santelices et al., 2017).

**Summary**

The literature on teacher quality reflects an abundance of research that has focused on teacher qualifications and practices, but little consensus exists around the extent to which each of these measures of teacher quality relates to student outcomes (Goe, 2007). Moreover, while numerous studies examining the relationships between individual components of teacher quality and student outcomes have been conducted, there remains a need for research that examines the extent to which components of teacher quality relate to one another (Santelices et al., 2017). Such research is relevant to gauging the construct validity of teacher quality within large-scale teacher evaluation policies that link measures of teacher practice, collected through classroom observations.
with other measures of teacher quality such as VAMs (Gregory, 2007; Santelices et al., 2017).

The role of context on measures of teacher quality is another factor that recent scholarship suggests may be important to consider in the development and implementation of evaluation schemes that comprise large-scale educational reform efforts. Although prior research has shown that differential patterns of teacher quality based on NCLB qualifications exist between various school contexts (Adamson & Darling-Hammond, 2011; O’Day & Smith, 2016), other findings suggest that this does not necessarily mean that teaching quality is better in some contexts and worse in others (Torres, 2018). More research is needed to understand the extent to which teacher practices mediate the relationships between teacher qualifications and student outcomes and whether any measure of teacher quality can be deemed a “context free” estimate of teacher quality (Santelices et al., 2017).
CHAPTER III

METHOD

The purpose of this study was to investigate the relationships between various measures of teacher quality and the influence of school district context on those measures. More specifically, this study aimed to analyze the extent to which teacher practices, as captured by evaluation ratings under SB 10-191, mediate the relationships between teacher qualifications and SB 10-191 teacher effect estimates, while investigating whether a school district’s size and geographical context moderates the relationships between these measures of teacher quality. While prior research has addressed questions about teacher qualifications and practices as separate measures of teacher quality in relation to student outcomes, few studies have examined these measures in tandem, or the extent to which teaching practices mediate the relationship between teacher qualifications and teacher effect estimates. Moreover, the literature is only beginning to address the role of school district context on measures of teacher quality under large-scale teacher evaluation systems. While initial findings from this research have suggested that no measures of teacher quality can be considered “context free,” there are no studies at this time that have directly investigated the extent to which school district contexts moderate the relationships between measures of teacher quality under such evaluation systems.

Using data gathered by CDE as part of its annual human resources data collection cycle, a quantitative quasi-experimental causal-comparative (ex-post facto) design (Fraenkel & Wallen, 2009; Salkind, 2010) was used to investigate the relationships
between teacher qualifications, teacher practices, and teacher effect estimates under SB 10-191. To account for the clustering of teachers in unique school district contexts, Multilevel Modeling (MLM) was used to separate teacher-level and school-level predictor variables in the analysis.

**Research Questions**

The research questions used in this study area as follows:

*Research Question 1:* To what extent are teacher qualifications associated with teacher effect estimates under SB 10-191?

*Research Question 2:* To what extent do teacher practices mediate the relationship between teacher qualifications and teacher effects?

2a. To what extent are teacher qualifications associated with better performance on the teacher quality standards used to measure practices under SB 10-191?

2b. After controlling for teacher qualifications, what is the relationship between teacher practice ratings and teacher effect estimates under SB 10-191?

2c. What is the relationship between teacher qualifications and teacher effect estimates under SB 10-191 after controlling for professional practice ratings?

*Research Question 3:* To what extent does school district context moderate the relationships between teacher qualifications and SB 10-191 teacher quality measures?
Research Design

Overview of Research Design

A quantitative quasi-experimental causal-comparative (ex-post facto) design was used to examine the relationships between various measures of teacher quality and school district context as reflected in statewide data collected by the CDE. This design was deemed most appropriate given that the purpose of causal-comparative research is to investigate the relationships between independent and dependent variables after an action or event has already occurred, while determining whether the independent variable affected the outcome of the dependent variable (Fraenkel & Wallen, 2009; Salkind, 2010). Such a design is also best suited to research in which manipulation of the independent variable is impossible or would otherwise be considered unethical (Salkind, 2010). In the case of SB 10-191 evaluation ratings, the data analyzed not only reflects evaluations of teachers that have already occurred, but also the grouping of teachers in to school district geographical classifications (contexts) existed prior to the research and reflects an independent variable that cannot be feasibly manipulated through an experimental design.

To account for the grouping of teachers (i.e. “clustering”) within unique school district contexts, Multilevel Modeling (MLM) was used as part of the statistical method for addressing the research questions. MLM is a complex form of regression analysis, which has become increasingly common for research in the education sector where data is often clustered within classrooms, schools, and district levels (Osborne, 2000; Raudenbush & Bryk, 2002; Woltman et al., 2012). By allowing researchers to analyze
variance in outcome variables in cases where predictor variables exist at different levels or hierarchies (i.e. teacher qualifications at the teacher level and school contexts at the school level), MLM has been found to contribute to more accurate estimates of shared variance within leveled or hierarchically structured data (Woltman et al., 2012). Given the nature of the research questions, which consider the influence of school-level predictors on teacher level predictors and outcomes, MLM was deemed an appropriate way to approach the statistical method for this study.

**Moderated Mediation Analysis**

Scholars interested in exploring relationships between variables must often account for situations in which the link between independent and dependent variables is not entirely direct, but rather is influenced by additional variables that help to explain the relationship between independent and dependent variables (Vogt, 2005). This is particularly true in the social sciences, where the complexity of living organisms and their environments give rise to “entities or processes that intervene” between inputs and outputs (Baron & Kenny, 1986, p. 1176). To account for such complexities in quantitative research, theorists have developed several methods for analyzing the effects of intervening variables that are hypothesized to influence the relationships between independent and dependent variables. These methods include *mediation analysis*, which examines the influence of intervening variables as transmitting agents between independent and dependent variables, and *moderation analysis*, which examines the influence of intervening variables on the direction and/or strength of the relationships between independent and dependent variables (Baron & Kenny, 1986; Bauer et al., 2006;
MacKinnon et al., 2000). Put another way, mediating variables can help to explain *how* the relationships between independent and dependent variables work whereas moderating variables can help to explain *under what conditions* such relationships will hold. In cases where both mediation and moderation are hypothesized to occur, the two methods can be combined via a single *moderated mediation model* that enables researchers to analyze the effects of mediating and moderating variables simultaneously, thereby increasing the reliability of the findings (Bauer et al., 2006; James & Brett 1984; Preacher et al., 2007).

Within the context of this study, moderated mediation analysis was deemed most appropriate for addressing the research questions which are concerned with analyzing the extent to which various teacher qualifications predict better performance on teacher effect estimates under SB 10-191 (*Research Question 1*); the extent to which teacher practices serve as transmitters (i.e. *mediate*) between teacher qualifications and teacher effect estimates (*Research Question 2*), and the role of school contexts in determining the strength (i.e. *moderating*) those relationships (*Research Question 3*). Figure 5 provides an illustration of the model used to address the research questions in this study through moderated mediation analysis.

In Figure 5, paths $a_1, b, c, \text{ and } c_1'$ form the basis of the mediation analysis that can be used to address *Research Questions 1 and 2*. 
Figure 5

Analytical Model for Moderated Mediation of Teacher Qualifications, Teacher Practices, and Teacher Effect Estimates

Path $a_1$ represents the effect of teacher qualifications on teacher practice ratings (with each specific qualification being measured by its own $a$ path in the analysis); path $b$ represents the effect of teacher practices (the qualifications on teacher practice ratings (with each specific qualification being measured by its own $a$ path in the analysis); path $b$ represents the effect of teacher practices (the hypothesized mediators) on teacher effect estimates; path $c$ represents the unmediated main effect (i.e. total effect) of teacher qualifications on teacher effect estimates; and path $c_1'$ represents the direct (unmediated) effect of teacher qualifications on teacher effect estimates, after controlling for teacher practices. Regression coefficients obtained from paths $a_1$ and $b$ were used to calculate the indirect effect (i.e. mediated effect) of teacher qualifications on teacher effect estimates and thus determine the extent to which teacher practices mediate the relationship between the two measures of teacher quality.
Paths $a_2$ and $c_2'$ represent the moderation (i.e. interaction) effects of school context on the relationship between teacher qualifications and teacher practices (path $a_1$), and teacher qualifications and teacher effect estimates (path $c_1'$), respectively. The regression coefficients obtained from these paths were used to address Research Question 3 while determining the extent to which moderated mediation has occurred (Hayes, 2015). Given that the measures for teacher practices and measures of student learning are collected within the same school through teacher evaluation under SB 10-191, these measures are already assumed to be conditioned by the school context. Thus, the moderating effects of school context examined by the study were restricted to paths $a_1$ and $c_1'$ of the mediation model.

**Theoretical Justification for Analytical Model**

The theoretical justification for the analytical model presented in Figure 5 is grounded in the literature on teacher quality, which includes some findings showing significant correlations (albeit of varying strength) between teacher qualifications and student outcomes (Goe, 2007). Specifically, a teacher’s education level, license type, and years of experience have been identified as variables of interest, as they have been most consistently cited as important characteristics for guiding policy decisions on teacher quality (Rice, 2003), yet little consensus has been achieved on how they affect student achievement (Goe, 2007).

Yet, relying on teacher’s “paper qualifications” alone has been deemed insufficient for assessing the quality of teachers (Barnett & Amrein-Beardsly, 2011; Goe, 2007). Therefore, information on teacher practices, which have also been found to be
correlated to student outcomes (Borman & Kimball, 2005; Danielson, 2006, 2007; Heneman et al., 2006; Holtzapple, 2003; Kimball et al., 2004; Milanowski, 2004), have been included in the model. By virtue of occurring on a consistent basis and defining the teaching/learning process between teachers and students, teacher practices are hypothesized to mediate the relationship between teacher qualifications and student outcomes.

Finally, the inclusion of school context as a moderating variable was based on the findings of recent research which suggests that teacher quality, as measured by evaluation of teacher practices and value-added estimates, can vary significantly between school contexts, but does not always reflect the uneven distributions of teacher qualifications between schools (Garcia, 2019; Torres, 2018). Thus, school context, and specifically, a school district’s rurality, is hypothesized to condition the relationships between teacher qualifications, teacher practices, and student outcomes.

**Research Sample and Data Sources**

The sample for this study included all teachers that received evaluation ratings under the Colorado State Model Evaluation System during the 2016-2017 school year. This year of data was selected as it provides the most recent information on teacher quality in Colorado under the original structure of SB 10-191. Given the research questions do not involve analyzing changes in teacher quality over time, but rather, focus on the extent to which various measures of teacher quality relate and are influenced by school district context, one year of data was deemed sufficient for conducting the analysis.
The source of data was based on records collected by CDE as part of its annual staff interchange human resource collection cycle. Following IRB approvals, the files were requested from the unit of Educator Effectiveness at CDE (CDE, 2019c) through established processes within the state agency. All data was masked prior to being shared with the researcher in order to prevent sharing of personally identifiable information. This masking included the use of random teacher and school district identifiers to minimize the risks of identifying specific teachers, particularly in districts with small populations of teachers. The files were stored and analyzed on a password-protected computer that was only available to the researcher. In all cases, adherence to protocols and processes as outlined by CDE and the University of Colorado Colorado Springs (UCCS) IRB were followed.

It is important to note that the CDE staff interchange collection allows school districts to report teachers as “not yet evaluated” or “no score” on any given standard or for an overall rating (Colorado State Board of Education, 2019c). This reporting can occur for various reasons including medical leaves and circumstances where an evaluation is undergoing an appeal. For the purposes of this study, cases in which teachers were not evaluated or received no score were screened out of the files prior to analysis and only teachers that received actual performance ratings will be included in the research sample.
Description of Variables

Dependent Variable

The dependent variable for the multilevel moderated mediation analysis was teacher evaluation ratings for Quality Standard VI of the Colorado State Board of Education Rules for teacher evaluation under SB 10-191 (Colorado State Board of Education, 2019). Quality Standard VI reads as follows:

- Quality Standard VI: Teachers take responsibility for student academic growth (CDE, 2015).

This standard is used to estimate teacher effects on student learning as reflected by a collection of state and locally determined assessments, which, collectively are referred to as Measures of Student Learning (MSLs). The State Board rules further specify that such measures must include:

- A measure of individually attributed Student Academic Growth, meaning that outcomes on that measure are attributed to an individual licensed person;
- A measure of collectively-attributed Student Academic Growth, whether on a school-side basis or across grades or subjects, meaning that outcomes on that measure are attributed to at least two licensed personnel (e.g., measures included in the school performance framework, required pursuant to section 22-11-204, C. R. S.);
- When available, statewide summative assessment results; and
• For subjects with annual statewide summative assessment results available in two consecutive grades, results from the Colorado Growth Model (Colorado State Board of Education, 2019, p. 18)

Based on the extent to which teachers meet annual student growth targets in each of the MSLs, Quality Standard VI ratings are assigned on a four-point ordinal scale that includes the categories of *Much Less than expected; Less than Expected; Expected*, and *More than Expected.* (CDE, 2017). Thus, for the purposes of conducting the analysis for this study, this variable was treated as ordinal, with codes 1, 2, 3, and 4 reflecting the categories of possible performance on SB 10-191 Standard VI.

**Independent Variables**

The independent variables for this study were comprised of a set of teacher qualification measures based on education level, license type, and years of experience. Each of these variables is described in further detail below.

**Education Level**

The staff interchange profile collected by CDE includes data on the highest level of education completed by teachers. The categories include 12 levels of education ranging from “No high school diploma” to “Doctoral (Doctors) degree.” In order to facilitate the analysis for this study, the data for this variable was recoded into a dichotomous variable (using conventional dummy coding of 0 and 1) that reflects whether or not a teacher holds a graduate level degree, with teachers not holding a graduate degree serving as the reference group (coded as 0). This approach is in alignment with the research and policy on teacher education levels, which primarily
differentiates between teachers who hold/do not hold a graduate degree (Clotfelter et al., 2010; Obonyo et al., 2018).

License Type

State data on the license types of educators is coded into 16 categories ranging from substitute authorization to Professional (Master) Teacher License. In order to facilitate the analysis for this study, the data for this variable was recoded into a dichotomous variable (using conventional dummy coding of 0 and 1) that reflects alternative/traditional licensure pathways. The alternative category included such designations as adjunct teacher, alternative teacher, interim teacher, substitution authorization, and charter school waiver, whereas the traditional category included designations of initial teacher, teacher, professional teacher, and professional master teacher. This binary approach is in alignment with the research on teacher quality, which is primarily concerned about differences in traditional and alternative licensure pathways on teacher effectiveness and student outcomes (Clotfelter et al., 2010; Henry et al., 2014; Lincove et al., 2015).

Years of Experience

The state data on teacher quality includes two measures of prior Pre/K-12 teaching experience, one that reflects total years of experience in and outside of Colorado schools, and the other, which only includes years of experience in Colorado schools. Both measures are self-reported by teachers and verified at the school district level. For the purposes of this study, the former count, which reflects total years of experience in all schools was used and was treated as a continuous variable.
Mediating Variables

The hypothesized mediating variables for this study included teacher practices as represented by SB 10-191 evaluation ratings on Teacher Quality Standards I, II and III, which Colorado State Board rules define as follows:

- Quality Standard I: Teachers demonstrate mastery of and pedagogical expertise in the content they teach. The elementary teacher is an expert in literacy and mathematics and is knowledgeable in all other content that he or she teaches.
- Quality Standard II: Teachers establish a safe, inclusive and respectful environment for a diverse population of students.
- Quality Standard III: Teachers plan and deliver effective instruction and create an environment that facilitates learning for their students.

While the Colorado State Model Evaluation system includes additional practice measures based on Reflection and Leadership (Quality Standards IV and V), only ratings on Standards I-III were used for the analysis as these standards most closely reflect classroom practices that directly affect student learning as demonstrated in the literature (Borman & Kimball, 2005; Danielson, 2006, 2007; Heneman et al., 2006; Holtzapple, 2003; Milanowski, 2004). Evaluation ratings for each practice are assigned based on a five-point ordinal scale that includes the categories of Basic, Partially Proficient, Proficient, Accomplished, and Exemplary.

To facilitate analysis of the impact of these practices on mediating the relationship between teacher qualifications and teacher effect estimates, the evaluation ratings for Quality Standards I, II and III, were collapsed into a single composite rating the reflects...
the combined effect of all practices. In order to preserve the original scale used to evaluate individual practices, a “preponderance of evidence” method, similar to the one used by CDE in implementing the evaluation system (CDE, 2019a) was used to round the average of the combined ratings to the nearest whole number. This new variable was treated as ordinal with codes 1, 2, 3, 4, and 5 reflecting categories of Basic, Partially Proficient, Proficient, Accomplished, and Exemplary, respectively. Teachers scoring in the Basic category will be used as the reference group for the purposes of the analysis.

**Moderating Variable**

The hypothesized moderating variable for this study was the rural classification of the schools in which teachers received their evaluation ratings during the 2016-2017 school year. CDE assigns a classification of rural to any schools that reside within districts that enroll less than 6500 students and small rural to any school that enrolls less than 1000 students (CDE, 2013b). Geographic location, in terms of distance from metropolitan areas is also used in determining the rural designation of Colorado schools (CDE, 2013b). In order to mitigate the risk of identifying teachers and schools in the small rural category, the designations small rural and rural were collapsed into a single designation of rural. To differentiate between non-rural and rural school district contexts in the analysis, a dichotomous variable (using conventional dummy coding of 0 and 1) was used with rural school districts, coded as 0, serving as the reference group.

**Statistical Method**

This study employed a multilevel moderated mediation analysis using regression to address the following research questions:
Research Question 1: To what extent are teacher qualifications associated with teacher effect estimates under SB 10-191?

Research Question 2: To what extent do teacher practices mediate the relationship between teacher qualifications and teacher effects?

2a. To what extent are teacher qualifications associated with better performance on the teacher quality standards used to measure practices under SB 10-191?

2b. After controlling for teacher qualifications, what is the relationship between teacher practice ratings and teacher effect estimates under SB 10-191?

2c. What is the relationship between teacher qualifications and teacher effect estimates under SB 10-191 after controlling for professional practice ratings?

Research Question 3: To what extent does school district context moderate the relationships between teacher qualifications and SB 10-191 teacher quality measures?

Regression Technique

Logistic regression has been well established in the methodological literature as a statistical method for analyses in which the dependent variable is not continuous (Iacobucci, 2012; Osborne, 2015). Still, there at least two types of logistic regression models (general and ordinal) that can be used to regress ordinal dependent variables and each has different implications for how the results of the analysis can be interpreted.
Whereas general logistic regression provides estimates that account for the grouped nature of ordinal data, it does not directly account for the ranking of groups based on categories (Osborne, 2015). Ordinal logistic regression provides a means of preserving the ranks of categories, but relies on the proportional odds assumption (also referred to as the parallel lines assumption), which states that the slope coefficients in the model are the same across all categories (Menard, 2001). This assumption is not always met when working with ordinal data and should be tested prior to performing analysis using ordinal logistic regression (Osborne, 2015).

To determine the most appropriate regression technique for this study, a series of full likelihood ratio tests comparing the fit of proportional odds models to models with varying location parameters was conducted for the dependent and mediating variables prior to analysis. These tests revealed that the data did not violate the proportional odds assumption, thus making ordinal logistic regression the most appropriate technique for this study.

**Model Specifications**

This study applied multilevel moderated mediation analysis using ordinal logistic regression to address the research questions. Figure 6 provides an illustration of the multilevel moderated mediation model with paths and variable labels. In single-level moderation analysis, the effect of the moderating variable (Z) and the predictor variable (X) on the dependent variable (Y) can be expressed through a basic multiple regression equation that includes an interaction term:

\[ Y = i + \beta_1 X + \beta_2 Z + \beta_3 XZ + e \]
where $i$ is the intercept, $\beta_1$ is the coefficient relating to the independent variable, $X$, $\beta_2$ is the coefficient relating to the moderator variable, $Z$, and $e$ is the residual (Fairchild & MacKinnon, 2009). Such an approach, if taken when the moderating variable lies at the group or “cluster,” level (i.e. Level-2) creates a biased estimate of the effects of the model by violating the assumption of independence that forms the basis of regression models (Bauer, et al., 2006; Hox, 2002; Kraft & de Leeuw, 1998; Raudenbush,
& Bryk, 2002). To account for this issue, a series of multilevel regression equations was composed that regressed each of the dependent variables on the independent variables following the specified paths, while accounting for the role of the Level-2 moderating variable on those paths. For path $a_1$, the model was specified as:

Level 1: $\ln(PRAC)_{ij} = \beta_{0j} + \beta_{1j}(EDUC)_{ij} + e_{ij}$

Level 2: $\beta_{0j} = \gamma_{00} + \gamma_{01}(RURAL)_{j} + u_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11}(RURAL)_{j} + u_{1j}$

where in the Level 1 equation $\ln(PRAC)_{ij}$ represents the ordered log of the odds of receiving an above basic rating on SB 10-191 professional practices for an individual teacher $i$ in school district $j$, $\beta_{0j}$ is the intercept term representing the average log odds of scoring above basic on SB 10-191 for teachers in school district $j$, $\beta_{1j}$ represents the effects of education level on the log odds of scoring above basic on SB 10-191 professional practices. The Level 2 equations provide estimates of the moderation effects of school context, represented by $\gamma_{01}, \gamma_{11}$, on the relationship between teacher education level and SB 10-191 professional practices ratings by treating the Level 1 intercept and predictor variable as dependent variable and regressing them on the school district’s rurality $(RURAL)$ status. $\gamma_{00}$ represents the estimated grand mean log odds of scoring above basic on SB 10-191 professional practices.

Similar models were used to estimate the effects of other teacher qualifications on SB 10-191 professional practice ratings, while measuring the influence of rurality as a moderating variable. For estimating the effects of license type $(LIC)$, path $a_2$, the model was expressed as:
Level 1: \( \ln(\text{PRAC})_{ij} = \beta_{2j} + \beta_{3j}(\text{LIC})_{ij} + e_{ij} \)

Level 2: \( \beta_{2j} = \gamma_{20} + \gamma_{21}(\text{RURAL})_{j} + u_{2j} \)

\( \beta_{3j} = \gamma_{30} + \gamma_{31}(\text{RURAL})_{j} + u_{3j} \)

For estimating the effects of experience \((\text{EXP})\), path \( a_i \), the model was expressed as:

Level 1: \( \ln(\text{PRAC})_{ij} = \beta_{4j} + \beta_{5j}(\text{EXP})_{ij} + e_{ij} \)

Level 2: \( \beta_{4j} = \gamma_{40} + \gamma_{41}(\text{RURAL})_{j} + u_{4j} \)

\( \beta_{5j} = \gamma_{50} + \gamma_{51}(\text{RURAL})_{j} + u_{5j} \)

C paths were also estimated using a series of models with \( \ln(\text{MSL})_{ij} \) being the Level 1 dependent variable representing the ordered log of the odds of receiving expected or higher rating on the SB 10-191 teacher effect measure. For estimating the effects of teacher education level \((\text{EDUC})\), the model was specified as:

Level 1: \( \ln(\text{MSL})_{ij} = \beta_{6j} + \beta_{7j}(\text{EDUC})_{ij} + e_{ij} \)

Level 2: \( \beta_{6j} = \gamma_{60} + \gamma_{61}(\text{RURAL})_{j} + u_{6j} \)

\( \beta_{7j} = \gamma_{70} + \gamma_{71}(\text{RURAL})_{j} + u_{7j} \)

For estimating the effects of license type \((\text{LIC})\), the model was specified as:

Level 1: \( \ln(\text{MSL})_{ij} = \beta_{8j} + \beta_{9j}(\text{LIC})_{ij} + e_{ij} \)

Level 2: \( \beta_{8j} = \gamma_{80} + \gamma_{81}(\text{RURAL})_{j} + u_{8j} \)

\( \beta_{9j} = \gamma_{90} + \gamma_{91}(\text{RURAL})_{j} + u_{9j} \)

And for estimating the effects of years of experience \((\text{EXP})\), the model was specified as:

Level 1: \( \ln(\text{MSL})_{ij} = \beta_{10j} + \beta_{11j}(\text{EXP})_{ij} + e_{ij} \)

Level 2: \( \beta_{10j} = \gamma_{100} + \gamma_{101}(\text{RURAL})_{j} + u_{10j} \)
\[ \beta_{11j} = \gamma_{110} + \gamma_{111}(RURAL)_j + u_{11j} \]

The \( b \) and \( c' \) paths were used to estimate the effects of each teacher qualification measure (education level (EDUC), license type (LIC), and experience (EXP)) on teacher effect estimates (MSL) as mediated by teacher professional practices (PRAC). Since the \( b \) and \( c' \) paths do not include a Level-2 moderator, each model was expressed as a single Level-1 equation. For estimating the mediated effects of education level (EDUC), the model was expressed as:

Level 1: \( \ln(MSL)_{ij} = \beta_{12j} + \beta_{13j}(PRAC)_{ij} + \beta_{14j}(EDUC)_{ij} + e_{ij} \)

Where \( \beta_{14j} \) represents the effects of a teacher’s education level (EDUC) on the log odds of receiving an expected or higher rating on the SB 10-191 teacher effect measure (MSL) (c’ path) as mediated by their professional practices (PRAC), which is estimated by \( \beta_{13j} \), (b path).

Similarly, the effects of licensure type (LIC) were estimated by the following equation:

Level 1: \( \ln(MSL)_{ij} = \beta_{15j} + \beta_{16j}(PRAC)_{ij} + \beta_{17j}(LIC)_{ij} + e_{ij} \)

And the effects of experience (EXP) were estimated as follows:

Level 1: \( \ln(MSL)_{ij} = \beta_{18j} + \beta_{19j}(PRAC)_{ij} + \beta_{20j}(EXP)_{ij} + e_{ij} \)

As a final step in the analysis, regression coefficients obtained from \( a \) and \( b \) paths were used to calculate the indirect effects (i.e. mediated effect) of teacher qualifications on teacher effect estimates using the Sobel (1982) test of dividing the product of the coefficients by the pooled standard error, \( \sigma_{ab} \), as follows:
\[
\text{Sobel test} = \frac{ab}{\sigma_{ab}}
\]

**Limitations**

Limitations have been defined in the literature as potential challenges to a study that lie outside of the control of the researcher (Simon, 2011). In quantitative research, limitations have the potential to affect the validity and reliability of data, which in turn can affect the validity and reliability of the research findings. One key limitation for this study concerns the quality of data used for conducting the analysis. While CDE provides an online statewide platform for districts to use in conducting teacher evaluations, the agency does not have authority to gather data from that system automatically for statewide reporting purposes. Rather, CDE relies upon self-reported “Educator Effectiveness Metrics” with assurances from individual school districts through annual human resource data collections (Colorado State Board of Education, 2019). Such a data collection procedure leaves open the possibility of inaccurate reporting or non-reporting of evaluation ratings on any given standard or overall rating (reflected in a rating of “no score” for an educator), which can pose a threat to the reliability of data in a given collection cycle. In order to minimize the threats to reliability that could result from inaccurate state reporting, all data were screened for non-reported and missing values prior to analysis so that only teachers who received ratings on all standards and an overall rating were included in the analysis.

A second limitation of this study stems from the suppression of personally identifiable data within small rural school districts, which is unavoidable given the privacy guidelines used by CDE (CDE, 2019). While providing a safeguard against the
ethical and methodological risks associated with working with small sample sizes 
(Forstmeir et al., 2016), this suppression does limit the interpretability of the findings and 
only allows for comparisons to be drawn between school districts that enroll more than 
6500 students or less than 6500 students. Future research may be necessary to determine 
the extent to which “degrees of rural” matter in affecting the relationships between 
measures of teacher quality and student outcomes, while overcoming the challenges of 
working with small sample sizes in small rural contexts.

A third limitation of this study was that it relied on indirect measures of teacher 
effects using the SB 10-191 evaluation ratings assigned to teachers based on measures of 
student learning (MSLs) for Quality Standard VI of the State Model Evaluation System. 
While SB 10-191 requires that this measure be weighted as 50% of a teacher’s overall 
evaluation rating, the law does not stipulate (outside of requiring state assessments, when 
available) what assessments school districts use to measure student growth. Indeed, the 
tradition of local control in Colorado gives school districts significant latitude in selecting 
measures of student learning and this can affect the validity and reliability of the teacher 
effect measures. Moreover, the Colorado evaluation system does not reflect a true value-added model (VAM) in which student prior learning and demographic characteristics are 
statistically controlled as part of calculating a teacher effect estimate (Aaronson, Barrow, 
& Sanders, 2003; Amrein-Beardsley & Holloway, 2019). Despite these limitations, the 
teacher effect estimates used in SB 10-191 reflect an essential component of teacher 
evaluation in Colorado, with significant implications for teacher tenure and retention, and
thus continue to hold value in this analysis, particularly with regard to how they related to other measures of teacher quality and are impacted by school context.

A final limitation of this study concerns the extent to which the findings of the study can be generalized to other samples, populations, and contexts (Frey, 2018). The sample restriction to include only teachers in Colorado schools could bias the results and thus threaten the comparability of teacher quality measures to educational contexts in other states or political contexts. To control for threats to external validity, teacher qualification measures were selected based on those most commonly referenced in the literature and policy discussions across political and geographical contexts (Goe, 2007; Goldhaber, 2002). Nonetheless, while external validity should be an important consideration in educational research, the dynamic theory of educational effectiveness, which was used as the theoretical framework to guide this study’s design, posits that education is significantly complex and thus assumes that the generalizability of findings in educational effectiveness may be contextually limited (Creemers, Kyriakides, & Charalambous, 2018). This has been noted elsewhere in the literature where scholars have increasingly highlighted the “power of contexts” (Berliner, 2002) as essential components of educational phenomena, and thus, the findings of this study continue to hold value in informing the body of literature on teacher quality and educational effectiveness research.

**Delimitations**

Delimitations are defined as the characteristics of research that are under the control of the researcher and limit the scope of a study (Simon, 2011). The decision to
use a single year of teacher evaluation data represents one such delimitation that prohibits
analysis of how teacher quality changes over time or the extent to which teacher
constructs under the original version of SB 10-191 compare with those in the revised
version of SB 10-191 implemented in the 2017-2018 school year. Both of these analyzes
can form the basis of future research on teacher quality and be strengthened by the
availability of additional data from multiple years of implementation of the revised SB
10-191 evaluation system.

Another delimitation in this study concerns the collapsing of state data into the
proxy variables that were used to perform the analysis. While state data on teacher
evaluation includes information on ratings received for each individual quality measure,
it does not include a single “composite” score for all practices. Thus, a proxy variable
representing overall performance on Quality Standards I, II, and III was calculated by
taking the average of all three individual scores and using a “preponderance of evidence”
method of rounding to the nearest whole number to preserve the scale. The use of or
proxy (i.e. derived) variables has been well documented within literature on quantitative
research design and is a common practice in educational research where “raw data” is
often not conducive to direct analysis (Gott & Duggan, 2003). Still the use of proxy
variables limits the internal validly of a study by resulting in a loss of information and
thus affect interpretation of findings.
CHAPTER IV

RESULTS

This study was designed to analyze the extent to which teacher practices as reflected in SB 10-191 evaluation ratings mediate the relationship between teacher qualifications and teacher effect estimates, while gauging the extent to which school district context, characterized by size and geographic location (i.e. rurality) moderates relationships between measures of teacher quality. Using a quantitative quasi-experimental causal-comparative (ex-post facto) design with MLM, this study addressed the hypothesis that teacher qualifications (inputs) are related to performance on SB 10-191 teacher quality measures of practice (processes), which in turn mediate the relationships between teacher qualifications and teacher effect estimates (outputs). This study also addressed the hypothesis that relationships between teacher qualifications and measures of teacher quality under SB 10-191 are not context-free, and can be influenced by the rurality status of a school district.

Data Screening and Descriptive Statistics

Prior to analysis, data were screened for missing values on any of the variables relevant to this study. This resulted in an exclusion of 5,978 cases from the original data file. The final research sample was composed of a total of 22,824 teachers and 983 schools. Descriptive statistics for the research sample are presented in Table 1.
### Table 1

**Descriptive Statistics of Research Sample**

<table>
<thead>
<tr>
<th>Teacher Level Variables</th>
<th>Total N (%)</th>
<th>Non Rural</th>
<th>Rural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher Count</td>
<td>22824</td>
<td>17102</td>
<td>5722</td>
</tr>
<tr>
<td>Education Level</td>
<td>22824</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- No advanced degree</td>
<td>10253 (45)</td>
<td>7071</td>
<td>3182</td>
</tr>
<tr>
<td>- Advanced degree</td>
<td>12571 (55)</td>
<td>10031</td>
<td>2540</td>
</tr>
<tr>
<td>License Type</td>
<td>22494</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Traditional</td>
<td>22436 (99)</td>
<td>16854</td>
<td>5582</td>
</tr>
<tr>
<td>- Alternative</td>
<td>58 (1)</td>
<td>41</td>
<td>17</td>
</tr>
<tr>
<td>Experience</td>
<td>22824</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- &lt; 5 years</td>
<td>4731 (21)</td>
<td>3474</td>
<td>1257</td>
</tr>
<tr>
<td>- 5 to 10 years</td>
<td>4329 (19)</td>
<td>3228</td>
<td>1101</td>
</tr>
<tr>
<td>- &gt;10 years</td>
<td>13749 (60)</td>
<td>10392</td>
<td>3357</td>
</tr>
<tr>
<td>SB 10-191 Practice Rating</td>
<td>22824</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Basic</td>
<td>36 (1)</td>
<td>28</td>
<td>8</td>
</tr>
<tr>
<td>- Partially Proficient</td>
<td>681 (3)</td>
<td>506</td>
<td>175</td>
</tr>
<tr>
<td>- Proficient</td>
<td>10275 (44)</td>
<td>7767</td>
<td>2508</td>
</tr>
<tr>
<td>- Accomplished</td>
<td>10160 (45)</td>
<td>7627</td>
<td>2533</td>
</tr>
<tr>
<td>- Exemplary</td>
<td>1672 (7)</td>
<td>1174</td>
<td>498</td>
</tr>
<tr>
<td>SB 10-191 Standard VI Rating</td>
<td>22824</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Much less than expected</td>
<td>171 (1)</td>
<td>118</td>
<td>53</td>
</tr>
<tr>
<td>- Less than expected</td>
<td>1908 (8)</td>
<td>1279</td>
<td>629</td>
</tr>
<tr>
<td>- Expected</td>
<td>12548 (55)</td>
<td>9145</td>
<td>3403</td>
</tr>
<tr>
<td>- More than expected</td>
<td>8197 (36)</td>
<td>6560</td>
<td>1637</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>School Level Variables</th>
<th>Non Rural</th>
<th>Rural</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Context</td>
<td>983</td>
<td>580</td>
</tr>
</tbody>
</table>
As shown in Table 1, the data screening for teacher evaluation ratings of “no score” led to a significant reduction of cases based on licensure type as a total of 330 alternatively licensed teachers were found to have received a rating of “no score” on SB 10-191 teacher evaluations for 2016-2017. The descriptive statistics also reveal that a relatively low number of teachers received ratings in the lowest two categories on SB 10-191 professional practices and teacher effect estimates on Standard VI. These smaller group n-sizes can affect the overall reliability of ordinal regression and thus, caution must be used when interpreting the results of the analysis (MacKinnon et al., 1995).

Table 2 provides a cross tabulation of teacher characteristics based on the independent variables used in this study.

**Table 2**

*Cross Tabulation of Teacher Education Level and Years of Experience*

<table>
<thead>
<tr>
<th></th>
<th>&lt; 5 years (%)</th>
<th>5 to 10 years (%)</th>
<th>&gt;10 years (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Advanced Degree</td>
<td>4101 (18)</td>
<td>2563 (11)</td>
<td>3589 (16)</td>
</tr>
<tr>
<td>Advanced Degree</td>
<td>1886 (8)</td>
<td>2640 (12)</td>
<td>8045 (35)</td>
</tr>
</tbody>
</table>

These statistics reveal that the smallest proportion of teachers are those with and advanced degree and less than 5 years of experience (8% of the total sample) and the largest proportion of teachers being teachers who hold an advanced degree and have more than 10 years of experience (35% of the total sample). Given that the proportion of teachers holding an advanced degree with more than 10 years of experience is nearly double that of teachers with no advanced degree having more than 10 years of
experience, it is possible that collinearity may exist between the independent variables used in the study. As such, caution must be used when interpreting the results of the analysis and drawing conclusions about the relative effects of teacher education level and years of experience on the dependent variable.

**Results of Statistical Analysis**

Analyses were conducted in accordance with the moderated mediation model in Figure 7. The results of the moderated mediation analyses are reported in terms of how they address each of the research questions used in the design of this study.

**Figure 7**

*Statistical Model for Moderated Mediation of Teacher Qualifications, Teacher Practices, and Teacher Effect Estimates with Variable labels*
**Teacher Qualifications and SB 10-191 Teacher Effect Estimates**

As a first step in the mediated moderation analysis, the main effects of teacher qualifications on teacher effect estimates were calculated in accordance with Research Question 1: *To what extent are teacher qualifications associated with teacher effect estimates under SB 10-191.* The main effects of teacher qualifications on teacher effect estimates is reflected in the c path in the moderated mediation model and is essential to the analysis, as a significant relationship between these variables must be shown to exist if mediation through teacher practices is possible or likely to occur (Barron & Kenny, 1986). Separate analyses using ordinal logistic regression were conducted for each teacher qualification measure and the results are summarized in Table 3.

**Table 3**

*Parameter Estimates for Ordinal Logistic Regression of SB 10-191 Teacher Effect Estimates on Teacher Qualifications*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>Wald</th>
<th>df</th>
<th>Odds Ratio</th>
<th>Confidence Interval</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>EDUC</td>
<td>0.46</td>
<td>0.03</td>
<td>306.83</td>
<td>1</td>
<td>1.58</td>
<td>1.50 1.67</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>LIC</td>
<td>-0.01</td>
<td>0.25</td>
<td>0.01</td>
<td>1</td>
<td>0.99</td>
<td>0.60 1.63</td>
<td>0.965</td>
</tr>
<tr>
<td>EXP</td>
<td>0.34</td>
<td>0.02</td>
<td>432.66</td>
<td>1</td>
<td>1.40</td>
<td>1.36 1.45</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

The results of the ordinal logistic regression revealed that only two of the three predictor variables hypothesized to be related to teacher performance on SB 10-191 Standard VI (MSL)– education level (EDUC) and years of experience (EXP) – were statistically significant. Specifically, the results of the ordinal logistic regression revealed that the odds of teachers holding an advanced degree obtaining a higher rating on SB 10-
Standard VI were found to be 1.58 (95% CI, 1.50 to 1.67) times that of teachers not holding an advanced degree ($X^2(1) = 306.83, p < .001$).

Additionally, increased experience was found to be associated with a 1.40 fold (95% CI, 1.36 to 1.45) increase in the odds of obtaining a higher rating on SB 10-191 Standard VI overall ($X^2(1) = 432.66, p < .001$). Overall, the results of the ordinal regression suggest that the more education and experience a teacher had, the greater the probability that teacher had of earning a higher rating on SB 10-191 Standard VI.

Since SB 10-191 Standard VI was designed to measure the contributions that teachers make to student learning (i.e. teacher effects), these findings provide some evidence that holding an advanced degree and having more experience are positively associated with more student learning. Furthermore, by confirming that a significant relationship exists between teacher education levels (EDUC), years of experience (EXP), and teacher effect estimates, these findings support additional analysis to determine the extent to which teacher practices mediate the relationship between teacher qualifications and teacher effects.

Teacher license type (LIC) was found to have no statistically significant effect on teacher effects estimates as analyzed in the model. This finding is likely due to the relatively high numbers of alternative teachers in Colorado that received a code of “no evaluation rating” on the state human resource data collection. Additional research and alternative data gathering techniques will be necessary to measure the full effects of teacher license type on teacher practices and teacher effect estimates. It should be noted that because license type (LIC) was found to have no significant relationship to teacher
performance on SB 10-191 Standard VI, it was not included as a relevant teacher qualification measure in any of the subsequent moderated mediation analyses used in this study.

Mediation Analysis

Having established that a significant relationship exists between teacher qualifications and teacher effects as measured by SB 10-191 Standard VI (MSL), the second phase of analysis involved testing for mediation in accordance with Research Question 2: To what extent do teacher practices mediate the relationship between teacher qualifications and teacher effects? To conduct the mediation analysis, a series of ordinal logistic regressions were used. The first series of ordinal logistic regressions examined the relationships between teacher qualifications and SB 10-191 teacher practice ratings. These relationships form the basis of the a paths in the model and addressed Research Question 2a To what extent are teacher qualifications associated with better performance on the teacher quality standards used to measure practices under SB 10-191? The results of the a path analysis are summarized in Table 4.

Table 4

<table>
<thead>
<tr>
<th>Parameter Estimates for Mediation Analysis (a paths)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>EDUC</td>
</tr>
<tr>
<td>EXP</td>
</tr>
</tbody>
</table>

With regard to the a paths, the results of the ordinal logistic regression showed a significant positive relationship between teacher education level (EDUC), years of
experience (EXP), and SB 10-191 professional practice ratings (PRAC). Specifically, the results revealed that the odds of teachers holding an advanced degree obtaining a higher rating on SB 10-191 professional practices were found to be 1.73 (95% CI, 1.65 to 1.82) times that of teachers not holding an advanced degree ($X^2(1) = 452.86, p < .001$).

Additionally, each year of experience was found to be associated with a 1.77 fold (95% CI, 1.72 to 1.83) increase in the odds of obtaining a higher rating on SB 10-191 professional practices ($X^2(1) = 1223.96, p < .001$). These results suggest that the more education and experience a teacher had, the greater the probability that teacher had of earning a higher rating on SB 10-191 professional practices.

Following the a path analysis, a second series of ordinal logistic regressions was used to examine the relationship between SB 10-191 teacher practice ratings and teacher effect estimates while controlling for teacher qualifications. These relationships reflect the b and c’ paths of the model and addressed Research Questions 2b After controlling for teacher qualifications, what is the relationship between teacher practice ratings and teacher effect estimates under SB 10-191 and 2c What is the relationship between teacher qualifications and teacher effect estimates under SB 10-191 after controlling for professional practice ratings? Table 5 provides estimates of the b and c’ path analyses.

Table 5

<table>
<thead>
<tr>
<th>Parameter Estimates for Mediation Analysis (b and c’ paths)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variable</strong></td>
</tr>
<tr>
<td>PRAC</td>
</tr>
<tr>
<td>EDUC</td>
</tr>
<tr>
<td>PRAC</td>
</tr>
<tr>
<td>EXP</td>
</tr>
</tbody>
</table>
With regard to the b paths, the results of the ordinal logistic regression revealed a statistically significant relationship between teacher SB 10-191 professional practice ratings (PRAC) and teacher ratings on SB 10-191 Standard VI (teacher effect estimates) after controlling for education level (EDUC) and years of experience (EXP). Specifically, the results show that after controlling for education level, the odds of teachers who earned higher rating on SB 10-191 professional practices also earning a higher rating on SB 10-191 Standard VI were 2.07 times greater (CI, 1.98 to 2.15) than those of teachers who earned lower ratings on SB 10-191 professional practices ($X^2(1) = 1292.12, p < .001$). Likewise, the odds of teachers who earned higher rating on SB 10-191 professional practices also earning a higher rating on SB 10-191 Standard VI were 1.24 times greater than those of teachers to earned lower ratings on SB 10-191 professional practices after controlling for teacher experience. Overall, the findings of this part of the analysis provide evidence of significant alignment between SB 10-191 ratings reflecting teacher practices and SB 10-191 ratings reflecting teacher effects on student growth and achievement.

The c’ paths of the mediation model provide parameter estimates that reflect the relationship between teacher qualifications and SB 10-191 Standard VI ratings, after controlling for SB 10-191 professional practice ratings (the proposed mediators). These estimates are provided in Table 4 and reveal that significant (direct) relationships exist between teacher education level (EDUC), experience (EXP), and SB 10-191 Standard VI ratings, after controlling for professional practices (PRAC). Specifically, teacher education levels were found to have a positive association with Standard VI ratings
where the odds of teachers with advanced degrees obtaining higher ratings on SB Standard VI were 1.40 times higher (95% CI, 1.33 to 1.48) than teachers not holding an advance degree after controlling for professional practices ($X^2 (1) = 160.46, p < .001$).

Similarly, having more years of experience contributed to the odds of a teacher obtaining a higher rating on SB 10-191 Standard VI by factor of 1.24 (95% CI, 1.20 to 1.28) over teachers with less experience after controlling for professional practices ($X^2 (1) = 161.77, p < .001$).

The findings of the mediation analysis suggest that partial mediation of teacher qualifications on teacher effects occurs through teacher practices (reflected in SB 10-191 professional practice ratings), as both the a path and b path coefficients were found to be statistically significant while the direct effect of teacher qualifications on teacher effects estimated by the c’ paths remained significant after controlling for the mediator (Barron & Kenny, 1986).

As a final step in the mediation analysis, and to determine whether the overall mediation effects were statistically significant, the regression coefficients obtained from the a and b paths were used to calculate the indirect effects (i.e. mediated effect) of teacher qualifications on teacher effect estimates using the Sobel (1982) test. The proportion mediated was also estimated using the method proposed by MacKinnon et al. (1995) for gauging the magnitude of indirect effects measured through logistic regression with the following formula:

$$\frac{ab}{(c' + ab)}$$
Table 6 provides a summary of the estimated indirect effects obtained from the Sobel test as well as the proportion mediated calculations.

**Table 6**

_Indirect Effects of Teacher Qualifications on SB 10-191 Teacher Effect Estimates_

<table>
<thead>
<tr>
<th>Indirect Effect</th>
<th>ab product</th>
<th>95% Confidence Interval</th>
<th>p value</th>
<th>Proportion Mediated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education Level (EDUC) through Professional Practices (PRAC)</td>
<td>0.41</td>
<td>0.09</td>
<td>0.71</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Experience (EXP) through Professional Practices (PRAC)</td>
<td>0.40</td>
<td>0.14</td>
<td>0.66</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

The results showed that professional practices were statistically significant mediators of both teacher education level and experience on teacher effect estimates. Moreover, the results of the proportion mediated calculations indicated that the magnitude of the indirect effects of teacher qualifications on teacher effect estimates was found to be greater for experience (PM = 0.66) than for education level (PM = 0.55). These findings validate the hypothesis (illustrated in Figure 7) that teacher qualifications (inputs) are related to performance on SB 10-191 teacher quality measures of practice (processes), which in turn mediate the relationships between teacher qualifications and SB 10-191 measures of student learning/teacher effect estimates (outputs). The findings also provide evidence that the overall magnitude of professional practices as mediators of teacher qualifications on teacher effects varies across teacher qualification measures.
Moderation Analysis

The final phase of this study involved analyzing the extent to which school district contexts, as characterized by school size and geographic location, (i.e. rural status) moderate the relationships between teacher quality measures, teacher practices, and teacher effect estimates as illustrated in the model (Figure 7). This analysis was performed in accordance with Research Question 3 To what extent does school district context moderate the relationships between teacher qualifications and SB 10-191 teacher quality measures and tested the hypotheses proposed by recent research (Santelices et al., 2017) that no measures of teacher quality, whether assessed by qualifications, practices, or teacher effects, can be interpreted as “context-free.” Multilevel ordinal logistic regression was used to estimate the moderation (also known as interaction) effects of rural status (RURAL) on the relationships between teacher qualifications and SB 10-191 professional practice ratings (a paths), as well as the c paths of the moderated mediation model. The results of the moderation analysis are summarized in Table 7.

Table 7

Interaction Effects of Rural Status on Teacher Qualifications, SB 10-191 Professional Practices, and SB 10-191 Teacher Effect Estimates

<table>
<thead>
<tr>
<th>Interaction Effect</th>
<th>Coefficient*</th>
<th>SE</th>
<th>Wald</th>
<th>df</th>
<th>Odds Ratio</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>a paths</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDUC*RURAL</td>
<td>-0.59</td>
<td>0.50</td>
<td>5021</td>
<td>983</td>
<td>0.55</td>
<td>0.21 1.45</td>
</tr>
<tr>
<td>EXP*RURAL</td>
<td>-0.04</td>
<td>0.06</td>
<td>4965</td>
<td>983</td>
<td>0.96</td>
<td>0.86 1.10</td>
</tr>
<tr>
<td>c paths</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDUC*RURAL</td>
<td>0.13</td>
<td>0.61</td>
<td>693</td>
<td>983</td>
<td>1.14</td>
<td>0.34 3.82</td>
</tr>
<tr>
<td>EXP*RURAL</td>
<td>0.11</td>
<td>0.06</td>
<td>843.6</td>
<td>983</td>
<td>1.12</td>
<td>0.99 1.26</td>
</tr>
</tbody>
</table>

Note: *p <0.05
The results of the moderation analysis revealed no significant effect of school district rural status on the relationships between teacher qualifications, SB 19-191 teacher practice ratings, and SB 10-191 teacher effect estimates. This finding invalidates the hypothesis that the strength of the relationship between teacher quality measures is dependent on school district context as characterized by rural status.

**Summary**

The results of the moderated mediation analysis showed that, overall, statistically significant relationships exist between teacher qualifications, SB 10-191 teacher practice ratings and teacher effect estimates measured by SB 10-191 Standard VI. More specifically, teacher practices (as measured by SB 10-191 professional practice ratings) were found to partially mediate the relationship between teacher qualifications (as measured by education level and years of experience) and teacher effect estimates (as measured by SB 10-191 Standard VI). Moreover, school district context, as characterized by size and geographic location (rurality), was found to have no significant moderating effects on the relationships between teacher education level and performance on SB 10-191 professional practice ratings and SB 10-191 teacher effect estimates.

The following chapter provides further analysis of the findings of the moderated mediation analysis and discusses them in light of each of the research questions. The findings are also considered in terms of how they contribute to existing literature on teacher quality. Limitations of interpretation are discussed along with implications for further research.
CHAPTER V
DISCUSSION

The purpose of this study was to investigate the relationships between various measures of teacher quality and the effects of school district context on those relationships. Using a dataset obtained from CDE, the study employed a quantitative quasi-experimental causal-comparative (ex-post facto) design (Fraenkel & Wallen, 2009; Salkind 2010) using ordinal logistic regression and multi-level modeling to analyze the extent to which teacher practices (as measured by SB 10-191 professional practice ratings) mediate the relationship between teacher qualifications and teacher effect estimates (as measured by SB 10-191 Standard VI). The analysis also examined the extent to which school district context, as characterized by size and geographic location (rurality), moderated the relationships between measures of teacher quality.

Summary of Conclusions

The findings of the moderated mediation analysis revealed that, in general, a teacher’s education level and years of experience are positively associated with teacher performance on the SB 10-191 professional practice measures, as well and teacher effect estimates. In addition, teacher practices were found to be statistically significant mediators of the relationship between teacher qualifications and teacher effect estimates, albeit, with varying magnitudes of mediation based on the specific teacher qualification variable analyzed. The magnitude of professional practices as a mediator of teacher experience was found to be higher than that of teacher education level.
In addition to the mediation findings, the findings of the moderation analysis revealed that school district context, as characterized by size and geographic location (rurality), is not a significant moderator of the relationships between teacher qualifications, teacher practices, and teacher effects. Together, the significant findings of the mediation analysis, along with the null findings of the moderation analysis, provide information that contributes to the emerging body of literature on teacher quality and school context, while informing discussions around the validity of teacher quality measures within high stakes teacher evaluation policies like SB-10-191. The findings also highlight the need for future research on teacher quality that focuses on investigating the specific mechanisms by which teacher practices (processes) mediate the relationships between teacher qualifications (inputs) and teacher effects (outputs).

Conclusions on Research Question 1

The relationship between teacher qualifications and teacher effects was investigated with Research Question 1 *To what extent are teacher qualifications associated with teacher effect estimates under SB 10-191?* The results showed both teacher education level and experience to be positively associated with performance on SB 10-191 Standard VI. The results also showed no significant relationship between teacher license type and teacher performance on SB 10-191. Each of these findings are further discussed below with regard to how they inform the existing literature on teacher quality.
Teacher Education Levels

With regard to the relationships between teacher education level and teacher effects, the findings of this study provide insights into what has been a largely inconclusive area of research on teacher quality. Indeed, while some research has shown only marginal effects of teachers holding advanced degrees on student achievement gains (Betts et al., 2003; Clotfelter et al., 2006; Clotfelter et al., 2010; Obonyo et al., 2019), other research has suggested that stronger correlations exist between teacher education levels and student achievement, particularly within science and math (Rice, 2003; Santelices et al., 2017).

Although the findings of this study are not enough to suggest a direct relationship between advanced degrees and student achievement gains, they do support the conclusion that, in general, teachers with advanced degrees have a greater likelihood of contributing to student achievement gains as measured by SB 10-191 Standard VI. As such, the findings of this study align with the findings of Rice (2003), Santelices et al. (2017) and others in highlighting teacher education level as an important indicator of teacher quality. The findings also provide support for what has been traditional school district human resource practice of awarding additional teacher compensation through salary schedules that rely on education level as one of the primary determinants of teacher quality. Such practices have been called into question by other research showing weak relationships between teacher education levels and student achievement (Goldhaber, 2002), but the findings of this study would suggest that teacher education level, as measured by whether
a teacher holds an advanced degree, can provide a valid measure of teacher quality for human resource purposes.

It is worth noting that one of the limits of the findings in this study is that they only reflect formal education in the form of holding an advanced degree, which is the only method used by CDE for measuring teacher education levels. Another area essential to understanding the effects of teacher education level on student achievement involves the professional development and training external to formal degree programs. Indeed, prior research has shown that professional development, and particularly that which aims to enhance teacher content knowledge and pedagogical content knowledge, has been linked to positive effects on student achievement (Harris & Sass, 2007; Newsom et al., 2019; Rice, 2003; Santelices et al., 2017). As such, there remains a need to better understand how teacher education that extend beyond formal degrees relate to student achievement and extent to which such education affects teacher effect estimates.

**Teacher Experience**

As with the findings on teacher education levels, the findings of this study around teacher experience provides insight into an area of educational research that has historically yielded mixed results. While some studies have provided evidence that teacher experience matters in terms of predicting student achievement (Betts et al., 2003; Hanushek & Rivkin, 2006; Rice, 2003), other studies have found no significant relationships to exist between years of experience and student achievement gains (Carr 2006; Gallagher, 2004). Still other studies have suggested that the positive effects teacher
experience on student achievement are confined to the first few years of teaching (Hanushek et al, 2005).

Given the findings of this study which showed a statistically significant relationship between teacher experience and SB 10-191 Standard VI ratings, it is logical to conclude that Colorado teachers who were more experienced had a greater likelihood of contributing to student achievement gains. As such, this finding adds important evidence to the existing research where the effects of teacher experience continue to be a matter of debate amongst scholars. Indeed, by providing contrasting evidence to the findings of Carr (2006) and Gallagher (2004) this study provides justification for greater scrutiny around studies showing no significant relationship between teacher experience and student achievement, while highlighting the need for continued research around the role of teacher experience. One possible reason for the difference in findings between this study and others may be that those conducted by Carr (2006) and Gallagher (2004) focused exclusively on teachers in charter schools, while the present study included teachers across all schools in Colorado that used the State Model Evaluation System (which includes some charter schools). As such, the findings of this study may be more relevant to addressing policy concerns in public schools such as awarding additional teacher compensation through salary schedules that rely on years of experience as key indicator of teacher quality. While such practices have been called into question by the research findings showing no link between teacher experience and student achievement, (Carr, 2006; Gallagher, 2004; Hanushek et al, 2005), the findings of this study provide
some evidence that years of experience should continue to be considered as a measure of assessing teacher quality and incentivizing teachers.

**Licensure Type**

The findings of this study revealed a null finding with regard to the relationship between teacher licensure type and SB 10-191 teacher effect estimates. This was likely due to the significant reduction in cases included in the research sample due to the relatively high numbers of alternatively licensed teachers that received a rating of “no score” on the Colorado State Model Evaluation System in 2016-2017. While Colorado law requires school districts to evaluate all licensed teachers on an annual basis (including those that are alternatively certified), the findings of this study suggest that many school districts may not be complying with that portion of the law.

In spite of the limitations presented by the data, this study does provide justification for further research around what factors may influence school district decision to not evaluate alternatively licensed teachers. Such research can not only shed additional light on how the experiences of alternatively licensed teachers differ from traditionally licensed teachers, but can also help to close gaps in the literature on how the distributions of teacher quality vary among teachers based on license type and school contexts (Fry & Anderson, 2011; Shuls & Ritter, 2013).

**Conclusions on Research Question 2**

Research Question 2 asked, *To what extent do teacher practices mediate the relationship between teacher qualifications and teacher effects?* The question was further divided into 2 parts, each representing a different phase of the mediation analysis. The
results of each part are discussed below in terms of how they address the research question.

**Teacher Qualifications and SB 10-191 Professional Practice Ratings**

Research Question 2a *To what extent are teacher qualifications associated with better performance on the teacher quality standards used to measure practices under SB 10-191* addressed the first part of the mediation analysis by investigating the relationship between teacher qualifications and SB 10-191 teacher practice ratings. The results showed that, overall, both teacher education levels and experience were positively associated with teacher ratings on SB 10-191 professional practices. Together, these findings add significant contributions to what has been a largely inconclusive area of educational research.

Broader interpretation of these findings requires careful attention to the outputs of ordinal logistic regression as well as the distributions of teachers across categories of the independent and dependent variables. On one hand, the findings could suggest that teacher education level and experience are significant contributors to higher qualities of teacher’s practice. This is reflected in the finding that increased levels of education and experience had a statistically significant effect on the odds of teachers earning SB 10-191 higher professional practice ratings compared to teachers with less experience. On the other hand, the fact that 60% of the research sample had more than 10 years of experience, combined with the fact that 96% of teachers in the research sample earned SB 10-191 professional practice ratings of *Proficient* or higher may have produced biased estimates around the relationships between teacher qualifications and teacher practices.
Such lack of overall distributions in overall teacher evaluation ratings continues to serve as a major point of criticism among detractors of large-scale teacher evaluation systems (Kraft & Gilmour, 2017; Schoales, 2017) and may limit the interpretability of research findings drawn solely from data collected by such systems.

In any case, the findings of this study help to advance the discourse around teacher qualifications and teacher effects in ways that are consistent with theoretical models on teacher quality. Indeed, despite such models suggesting a link between teacher qualifications and teacher practices (Goe, 2007), the majority of research on teacher quality has treated them as distinct domains and few studies have sought to address the extent to which teacher qualifications contribute to stronger practices. One recent study by Santelices et al. (2017) found no significant effect of teacher education level or years of experience on predicting teacher evaluation ratings used to measure teacher practices and teacher effects under Chile’s teacher evaluation system. The results of this study provide some contrasting evidence to Santelices et al. (2017), while suggesting that more research is necessary to understand the extent to which data gathered through large scale teacher evaluation system can be applied to examining the relationships between qualifications and teacher practices.

**SB 10-191 Professional Practice Ratings and Teacher Effect Estimates**

Research Question 2b *After controlling for teacher qualifications, what is the relationship between teacher practice ratings and teacher effect estimates under SB 10-191* was designed to investigate the relationships between teacher practices and teacher effects and the results showed that, after controlling for teacher education level and
experience, significant relationships continued to exist between teacher practice ratings and teacher effect estimates. These findings are not only important for confirming the results of previous studies that have found significant links between teacher practices and student achievement (Borman & Kimball, 2005; Heneman, Milanowski, Kimball, & Oddden, 2006; Holtzapple, 2003; Milanowski, 2004), they also help to address the growing demand for determining the extent to which different standards-based criteria used to measure teacher quality in large-scale evaluation systems align with one another (Santelices et al., 2017). Such alignment is referred to in the literature as the convergent validity of measures of teacher quality and is defined by Gregory (2007) as the extent to which two sets of similar constructs are correlated. While the present study did not directly measure the correlations between SB 10-191 professional practice ratings and SB 10-191 teacher effect estimates, the results of the mediation analysis suggest a positive association between the two constructs of teacher quality under the State Model Evaluation System. In general, teachers who demonstrated higher levels of practice had a greater likelihood of also demonstrating higher levels of student growth and achievement as measured by SB 10-191 Standard VI.

Given recent policy discussions that have called into question Colorado’s approach to measuring teacher quality through evaluation that is evenly based upon professional practice ratings and teacher effect estimates (Meltzer & Fish, 2019), the findings of the study hold relevance by providing evidence that the teacher quality measures captured by the evaluation system yield results that are consistent with what would be expected from what is theorized in the literature as a positive relationship
between teacher practices and teacher effects (Goe, 2007). Specifically, the findings of this study provide evidence, in addition to that already noted in the literature (Danielson, 2006, 2007) that higher quality practices are associated with larger teacher effects on student achievement. Moreover, the ratings assigned to teachers in the evaluation system under SB 10-191 reflect distributions of professional practices and teacher effect estimates that are correlated enough to allow this relationship to be detected through statistical analysis. In the absence of statistically significant results, the findings of the mediation analysis may have provided evidence in support of the contention that it is unfair to tie such a significant proportion (50%) of teacher evaluations to teacher effects (Metzler & Fish, 2019) as it would indeed be unfair if one measure of teacher quality (based on practices) yielded evaluation results that were different than another (based on teacher effects). The statistically significant results yielded by the mediation analysis, however, provide evidence that the teacher evaluation system under SB 10-191 is doing exactly what it was designed to do in terms of providing information on teacher quality using a combination of measures that are based on research findings that suggest strong relationships between teacher practices and teacher effects on student achievement (Danielson, 2006, 2007; Goe; 2007).

Beyond contributing to modern debates on teacher evaluation, the findings of the analysis reflecting a positive association between teacher practices and teacher effects also holds relevance to policy and practice around educator preparation by suggesting that a focus on developing the professional practices of emerging educators (and specifically, domains of practice that involve instructional design, classroom
management, and facilitation of learning) holds promise for contributing to student achievement, particularly as educators continue to develop their practices through experience. Such a conclusion aligns with findings in the literature which suggest that a focus on developing pedagogical practice is an important component of educator preparation (Darling-Hammond et al., 2005). Moreover, this conclusion provides some support for traditional teacher education pathways which have been found to have a stronger focus on developing pedagogical knowledge than alternative pathways to teaching (Darling-Hammond et al., 2005).

**Summary Conclusions on Mediation Analysis**

Taken together, the findings of Research Questions 2a and 2b can be applied to addressing Research Question 2 *To what extent do teacher practices mediate the relationship between teacher qualifications and teacher effects?* These findings revealed partial mediation of teacher qualifications on teacher effects occurs through teacher practices as all mediation path coefficients were found to be statistically significant while the direct relationship between teacher qualifications on teacher effects remained significant after controlling for the mediator (Barron & Kenny, 1986).

When interpreting these results, it is important to note that the finding of a significant indirect effect of teacher qualifications through teacher practices is evidence consistent with a mediational hypothesis, but does not in and of itself prove that a causal pathway is the reason for the association. Indeed, to suggest that more education and experience will always *cause* higher levels of teacher practice which it turn will *cause* gains in student achievement would be to ignore one of the critical components of the
dynamic theory of education effectiveness, which provided the theoretical basis for the
design of the study. Rather, the findings of this study suggest that more education and
experience is associated with higher levels of practice, which in turn is associated with
higher teacher effect estimates on student achievement. This distinction is important as it
assumes that other factors beyond those directly measured in this study may affect the
relationships between teacher qualifications, teacher practices and teacher effects. The
distinction also holds importance for how educational practitioners and policy makers
should apply the findings of this study as simply increasing the education levels and years
of experience are not in themselves guarantees of higher student achievement (Creemers
& Kyriakides, 2006).

Despite these limitations, the findings of the mediation analysis help to advance
theoretical understandings of teacher quality by validating the hypothesis proposed by
Goe (2007), which suggests that, rather than producing direct (linear) effects on student
achievement, teacher quality factors can best be conceived of comprising of various
inputs, processes, and outputs that ultimately affect student achievement. This model,
while rarely tested through empirical research that includes all components, has been
essential in the identification of teacher practices (processes) as one of the central
domains of teacher quality measured through teacher evaluation and other policies on
teacher quality. Given the findings of the mediation analysis which suggest that teacher
practices can serve as significant transmitters of the effects of teacher qualifications on
student achievement, this study provides support for the continued use of teacher
evaluation constructs and other mechanisms that emphasize professional practice as a
significant domain of teacher quality (Borman & Kimball, 2005; Danielson, 1996, 1997; Heneman et al., 2006; Holtzapple, 2003; Milanowski, 2004; Santelices et al., 2017).

Conclusions on Research Question 3

The effects of school context on measures of teacher quality was investigated by Research Question 3 To what extent does school context moderate the relationships between teacher qualifications and SB 10-191 teacher quality measures? The purpose of this investigation was to test the hypothesis that measures of teacher quality in Colorado would be moderated by the rural status of the school. The findings of the moderation analysis showed no significant interaction effects of school rural status on the relationships between teacher qualifications and SB 10-191 teacher practice ratings, nor on the relationships between teacher qualifications and teacher effect estimates measured by SB 10-191 Standard VI.

While these results invalidate the hypothesis that school rural status moderates the relationships between teacher qualifications and SB 10-191 measures of teacher quality, it is important to consider them in terms of how Colorado defines rural schools. Indeed, given that the definition of rural in Colorado is based on school size and geographic location, the findings of the moderation analysis can be interpreted as suggesting that school size and geographic location had no significant effect on the relationships observed in the mediation analysis. Put another way, this null finding provides support for a new hypothesis that the significant relationships observed in the mediation analysis do not change based on school size and geographic location across Colorado schools.
More research is necessary to test this hypothesis and determine the extent to which school size and geographic location affect teacher quality in Colorado.

Within the context of the extant literature on teacher quality, the findings of this study are in contrast with those of other studies examining the effects of rurality on measures of teacher quality (Isenberg et al., 2013; Sass et al., 2012; Torres, 2018). A likely reason for this difference may be in how ‘rural’ is defined in Colorado compared to other locales. Indeed, as Arnold et al. (2007) note, definitions of rural vary considerably across the country and while most definitions include size and geographic location as key components of rurality, some definitions include factors that go beyond size and geographic location and include such as mean socio-economic status (SES). Thus, how governments define rural has important implications for educational research, policy and resource distributions and a null finding with regard to the effects of rurality in Colorado is not enough to suggest that rurality has no impact on teacher quality measures. More research is necessary to determine whether factors associated with rurality in Colorado (such as mean SES), but not included in the definition of rurality play a significant role in moderating the relationships between teacher quality measures.

**Implications for Policy**

Modern policy discussions are replete with debates about teacher quality, the best ways to measure teacher quality, and the application of teacher quality measures to decisions around hiring, firing and compensation of teachers. Often at stake in these debates is the question of whether teacher quality can best be addressed through educational policy, initial teacher preparation, and/or ongoing professional development
that improves teacher content knowledge and professional practices (Santelices et al., 2017). The findings of this study provide support for policies that enhance teacher qualifications by incentivizing advanced degrees and gaining experience in the field of education. Indeed, given the positive associations observed between teacher education level, experience, and teacher effects, it is worth considering how schools can recruit additional teachers with advanced degrees and, perhaps more importantly, retain them in the teaching profession. Current efforts to incentivize additional education and years of experience through teacher salary schedules are considered valid given the findings of this study and should be sustained as practices that promote teacher quality across school contexts in Colorado.

With regard to policy debates on teacher evaluation, the findings of this study provide evidence in support of the convergent validity of the Colorado Model Evaluation System which appropriately emphasizes teacher practices and teacher effects as essential domains of teacher quality. Given the positive associations observed between teacher qualifications, teacher practice ratings, and teacher effect estimates, it is worth considering whether additional measures of teacher quality based on qualifications should be considered as part of the state evaluation model. Specifically, years of experience, having been observed to reflect a stronger magnitude of mediation through teacher practices, may warrant additional consideration as a valid measure of teacher quality that can supplement the measures of teacher quality (practices and teacher effects) already included in the Colorado State Model Evaluation System.
The findings of the moderation analysis suggest that Colorado policy makers should carefully consider how rurality is defined and applied within educational policy decisions and distributions of resources. Indeed, while Colorado does make a distinction between rural (1000-6000 students) and small rural schools (less than 1000 students), such a narrow definition of rural that is only based on considerations of school size and geographic location may not be enough to reflect the contextual differences that exist between rural and non-rural schools (CDE, 2013b). Bringing the definition in line with constructs that have be found to be significant in education policy and resource distribution by including such factors as mean SES (Frisby & Reynolds, 2005; Santelices et al., 2017) may yield more equitable policies that support teacher quality across varying school contexts.

Implications for Future Research

The findings of this study provide several justifications for future research on teacher quality and evaluation. Specifically, the findings of Research Question 1 To what extent are teacher qualifications associated with teacher effect estimates under SB 10-191 provide support for additional research that validates the findings around the effects of teacher education level and experience on student achievement. Such research should take into account the methods used across state and local contexts to gauge teacher effects (i.e. what specific assessments and data are included in teacher effect estimates and how it is weighted), as well as the modes of educating teachers through formal training and degree programs and professional development. The null findings related to Research Question 1 also provide an opportunity for research exploring school district
practices in evaluating alternatively licensed teachers as well as additional studies that examine the effects of licensure type on teacher quality.

With regard to Research Question 2 *To what extent are teacher qualifications associated with better performance on the teacher quality standards used to measure practices under SB 10-191* the findings of this study provide justification for future research that more closely examines the mechanisms by which teacher practices mediate the relationships between teacher qualifications and teacher effects. Indeed while it would be logical to conclude that more formal education contributes to higher quality teaching, more research is necessary to determine the extent to which type and focus of training (i.e. degree focus, content) matters in terms of how it affects professional practices. Likewise, additional research is necessary to validate the findings of teacher experience and confirm whether experience has a direct effect on teacher practices or whether the mean experience levels of the teacher population lead to biased estimates of the effects of teacher experience.

Finally, with regard to school district context and rurality examined by Research Question 3 *To what extent does school context moderate the relationships between teacher qualifications and SB 10-191 teacher quality measures* the null findings provide opportunities for research exploring the extent to which factors associated with rurality, but not included in Colorado’s definition of rurality, affect relationships between measures of teacher quality in Colorado. Of key interest is the question of whether the expansion of the definition of “rural” beyond size and geographic location to include additional factors such as mean SES, would lead to research findings more closely
aligned to the differences in teacher quality and student achievement known to exist between rural and non-rural schools (Adamson & Darling-Hammond, 2011; O’Day & Smith, 2016; Santelices et al., 2017).

**Concluding Summary**

The purpose of this study was to investigate the relationships between measures of teacher quality and the effects of school context on those measures. Such an investigation is relevant to modern debates on teacher quality and evaluation as well as policies that seek to ensure that all students have access to quality teachers. Moreover, recent research on teacher quality has posited that measures of teacher quality assessed through large scale evaluation systems may not yield results that are free of the contexts in which they are obtained (Santelices et al., 2017).

Lingering disagreements around what matters most in teacher quality has caused substantial shifts in the focus of federal and state policies aimed at increasing teacher quality and student learning. These shifts are most obvious in the transition from NCLB (2002) which treated teacher quality through the narrow lens of “paper qualifications” (Barnett & Amrein-Beardsly, 2011; Goe, 2007) to RTT (2011) and ESSA (2015), which both placed high stakes on teacher practices and student results. Frustrations have emerged around uneven implementation of teacher evaluation policies, causing some to call into question the value and fairness of large-scale evaluation systems for measuring teacher quality (Close et al., 2018; Meltzer & Fish, 2019; Ramirez et al., 2014). Still, research findings continue to demonstrate that teacher quality matters and that regular
supervision and feedback of teachers is important for improving the quality of teaching (Goe, 2007; Frontier & Mielke, 2016).

The findings of this study are important to informing these debates as they provide evidence that helps to address disagreements and close gaps in the theoretical and practical literature on teacher quality and evaluation. Specifically, they encourage the continued use of instruments to measure and incentivize teacher quality, as these instruments appear to be working in terms of capturing relevant data that is consistent with the findings of the literature (Goe 2007).

In the effort to address modern policy questions around whether large-scale teacher evaluation systems should be mended or ended, this study proposes that teacher quality measures reflecting teacher education levels, experiences, and practices remain a central component of school improvement and accountability through evaluation, as this has the potential of ensuring that all students have access to high quality teachers. Practitioners and policy makers may risk “throwing the baby out with the bath water” if the strengths of teacher evaluation systems are not acknowledged in terms of their ability to capture meaningful data around important and interrelated domains of teacher quality.
REFERENCES


APPENDIX A

IRB Approval Letter

University of Colorado
Colorado Springs

Institutional Review Board (IRB) for the Protection of Human Subjects

Date: 1/30/2019

IRB Review [APPROVED]

IRB PROTOCOL NO.: 19-099
Protocol Title: A Policy Implementation Study on Colorado’s Educator Effectiveness System under SB 10-191
Principal Investigator: Curtis Garcia
Faculty Advisor if Applicable: Sylvia Mendez
Application: New Application
Type of Review: Exempt Category 4
Risk Level: No more than Minimal Risk
Renewal Review Level (If changed from original approval) if Applicable: N/A No Change
This Protocol involves a Vulnerable Population: N/A (No Vulnerable Population)
Expires: N/A

*Note, if no expiration date is indicated: Changes in the research need to be approved before implementation, and you need to report any adverse events. Requests for status updates may be sent by the IRB. In addition, the protocol may match more than one review category not listed.
Externally funded: ☐ No ☑ Yes
OSP #: Sponsor:

Thank you for submitting your Request for IRB Review. The protocol identified above has been reviewed according to the policies of this institution and the provisions of applicable federal regulations. The review category is noted above, along with the expiration date, if applicable.

Once human participant research has been approved, it is the Principal Investigator’s (PI) responsibility to report any changes in research activity related to the project:
- The PI must submit all protocol, recruitment, advertising, and consent form amendments/revision to the IRB for approval.
- The IRB must approve these changes prior to implementation.
- If you are a student, please note that it is required to include the IRB approval letter to the library when you submit the dissertation/thesis.
- The PI must promptly inform the IRB of all unanticipated serious adverse events (within 24 hours). All unanticipated adverse events must be reported to the IRB within 1 week (see 45CFR46.108(b)(4)(ii)). Failure to comply with these federally mandated responsibilities may result in suspension or termination of the project.
- If required, renew the study with the IRB at least 10 business days prior to expiration.
- Notify the IRB when the study is complete

If you have any questions, please contact Research Compliance Program Director in the Office of Sponsored Programs and Research Integrity at 719-255-3903 or arb@uccs.edu

Thank you for your concern about human subject protection issues, and good luck with your research.

Sincerely yours,

Zek Valkyrie
Zek Valkyrie, PhD
IRB Reviewer

www.uccs.edu/oas/ossr
Version 1.1.12018

1420 Austin Bluffs Parkway Colorado Springs, CO 80918 719-255-3221 phone 719-255-3706 fax

117
Date: 02/19/2020

IRB PROTOCOL NO.: 19-099
Protocol Title: A Policy Implementation Study on Colorado’s Educator Effectiveness System under SB 10-191
Principal Investigator: Curtis Garcia
Faculty Advisor if Applicable: Robert Mitchell
Application: Request for Change
Type of Review: Exempt 4
Risk Level: Minimal
This Protocol involves a Vulnerable Population:
Expires: Exempt
*Note: If exempt, if there are no major changes in the research, protocol does not require review on a continuing basis by the IRB. In addition, the protocol may match more than one review category not listed.
Externally funded:
OSP #: Sponsor:

Thank you for submitting your Request for IRB Review to change population (the population is now limited to evaluation data for teachers in 2016-2017 only), add additional demographic measures (teacher education level, license type, and years of experience), add new faculty advisor, and updated study title. The protocol identified above has been reviewed according to the policies of this institution and the provisions of applicable federal regulations. The review category is noted above, along with the expiration date, if applicable.

Once human participant research has been approved, it is the Principal Investigator’s (PI) responsibility to report any changes in research activity related to the project:
- The PI must submit all protocol, recruitment, advertising, and consent form amendments/revisions to the IRB for approval.
  - The IRB must approve these changes prior to implementation.
- Changes in funding status must be reported to the IRB as quickly as possible to ensure funding requirements are met.
- If you are a student, please note that it is required to include the IRB approval letter to the library when you submit the dissertation/thesis.
- The PI must promptly inform the IRB of all unanticipated serious adverse events (within 24 hours). All unanticipated adverse events must be reported to the IRB within 1 week (see 45CFR46.103(b)(5)). Failure to comply with these federally mandated responsibilities may result in suspension or termination of the project.
- The PI must submit a Continuing Review/Renewal application to the IRB at least 10 business days prior to expiration to continue projects beyond the expiration date (if applicable).
- Notify the IRB when the study is complete.

If you have any questions, please contact Research Compliance Program Director in the Office of Sponsored Programs and Research Integrity at 719-255-3903 or info@uccs.edu.

Thank you for your concern about human subject protection issues, and good luck with your research.

Sincerely yours,

Carmen Stavroulou