

2022

## The Relationship of Coaching on Quality Ratings in Early Childhood Care

Michelle Renee Burkhard  
*Walden University*

Follow this and additional works at: <https://scholarworks.waldenu.edu/dissertations>



Part of the [Education Commons](#)

---

This Dissertation is brought to you for free and open access by the Walden Dissertations and Doctoral Studies Collection at ScholarWorks. It has been accepted for inclusion in Walden Dissertations and Doctoral Studies by an authorized administrator of ScholarWorks. For more information, please contact [ScholarWorks@waldenu.edu](mailto:ScholarWorks@waldenu.edu).

# Walden University

College of Education

This is to certify that the doctoral dissertation by

Michelle Burkhard

has been found to be complete and satisfactory in all respects, and that any and all revisions required by the review committee have been made.

## Review Committee

Dr. John Flohr, Committee Chairperson, Education Faculty  
Dr. Andrew Thomas, Committee Member, Education Faculty  
Dr. Charlotte Redden, University Reviewer, Education Faculty

Chief Academic Officer and Provost  
Sue Subocz, Ph.D.

Walden University  
2022

Abstract

The Relationship of Coaching on Quality Ratings in Early Childhood Care

by

Michelle Burkhard

MA, Central Michigan University, 2009

BS, Central Michigan University, 2002

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education

Walden University

February 2022

## Abstract

Early childhood educators serving children ages birth through 8 years seek to continuously improve the quality of the learning environment and curriculum practice. All 50 states created a Quality Rating Improvement System (QRIS) to assess childcare center program quality and provide coaching for teachers. Early childhood education programs with intentional coaching score higher on several measures of quality associated with better child outcomes. Research demonstrates that there is a gap in understanding the relationship of coaching to quality scores within one Northcentral state's childcare centers. The purpose of this quasi experimental quantitative study was to determine the difference in mean QRIS scores between two groups of childcare centers. Vygotsky's sociocultural theory of learning provided a theoretical framework. The research questions addressed to what extent there was a difference in initial quality rating scores upon QRIS entry and difference scores after prolonged time in the QRIS. Archival data were used from a population of 208 center- and home-based programs that were quality rated. One-way analysis of variance results showed that the pre-coaching group scored higher on the initial quality rating than programs without prior coaching. Analysis of covariance results showed the groups that previously received no coaching increased more from the initial to final rating relative to the pre-coaching group. The findings of the present study may help educators and parents who initiate and support early childhood systems analyze current supports and renegotiate financial supports to raise program quality and promote positive social change.

The Relationship of Coaching on Quality Ratings in Early Childhood Care

by

Michelle Burkhard

MA, Central Michigan University, 2009

BS, Central Michigan University, 2002

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education

Walden University

February 2022

## Dedication

I would like to dedicate this work to the incredibly resilient and hardworking early childhood professionals across the field who strive to humbly serve families, children, and their communities daily. I am incredibly grateful for the work you do and honored to call you friend and colleague.

## Acknowledgments

Anyone who has entered the Ph.D. arena knows a project like this is never a solo endeavor. While sometimes it seemed very lonely, I was never alone. I would like to acknowledge the people with the greatest impact, influence, and support that helped me finish this incredible journey.

First of all, I would like to thank my family for making me meals, giving me space, and yes, sometimes even parenting yourself for a bit. I couldn't have done it without you. I now receive the title Doctor but feel you should, too!

Second, I am eternally grateful to Dr. John Flohr and Dr. Andrew Thomas for the incredibly painstaking task of being my committee. I really appreciate the great conversations over the years, the getting to know you, and yes, even the ugly bits along the way. Thank you for investing in me!

Next, I'd like to highlight a special group of people that make up my mentors, coaches, fellow students, and colleagues. You've all encouraged, supported, and helped sustain me and I'm incredibly grateful.

Finally, I tip my hat to all the previous, present, and future researchers in the field of early childhood education. Keep up the good fight. The world needs our work as we are helping to form the future for children, families, and communities!

## Table of Contents

|  |    |
|--|----|
| List of Tables .....                             | iv |
| List of Figures .....                            | v  |
| Chapter 1: Introduction to the Study.....        | 1  |
| Background of the Study .....                    | 3  |
| Problem Statement .....                          | 5  |
| Purpose of the Study .....                       | 5  |
| Research Questions and Hypotheses .....          | 6  |
| Theoretical Framework.....                       | 7  |
| Nature of the Study .....                        | 8  |
| Operational Definitions.....                     | 9  |
| Assumptions.....                                 | 11 |
| Scope and Delimitations .....                    | 12 |
| Limitations .....                                | 12 |
| Significance of the Study .....                  | 12 |
| Summary.....                                     | 13 |
| Chapter 2: Literature Review.....                | 15 |
| Literature Search Strategy.....                  | 16 |
| Theoretical Framework.....                       | 16 |
| Literature Review Related to Key Variables ..... | 19 |
| History of the Problem.....                      | 19 |
| Constructs of Interest .....                     | 24 |



|   |    |
|---|----|
| One Northcentral State’s Quality Rating Improvement System..... | 36 |
| Methodology .....   | 38 |
| Justification of Variable Selection .....                       | 40 |
| Summary .....   | 60 |
| Chapter 3: Research Method.....                                 | 62 |
| Research Design and Rationale .....                             | 62 |
| Methodology .....   | 64 |
| Population .....  | 64 |
| Sampling and Sampling Procedures .....                          | 64 |
| Archival Data Procedures .....                                  | 65 |
| Instrumentation .....   | 66 |
| Operationalization of Variables .....                           | 67 |
| Data Analysis Plan .....  | 68 |
| Research Question and Hypotheses .....                          | 68 |
| Threats to Validity .....                                       | 70 |
| Ethical Procedures .....  | 71 |
| Summary .....   | 72 |
| Chapter 4: Results .....  | 73 |
| Data Collection .....   | 74 |
| Results   | 79 |
| Research Question 1 .....                                       | 80 |
| Research Question 2 .....                                       | 84 |

|   |     |
|---|-----|
| Summary .....   | 87  |
| Chapter 5: Discussion, Conclusions, and Recommendations .....                 | 89  |
| Interpretation of Findings to the Empirical Literature .....                  | 90  |
| Interpretation of Findings in Relationship to the Theoretical Framework ..... | 92  |
| Limitations of the Study.....   | 93  |
| Recommendations.....  | 94  |
| Implications for Positive Social Change.....                                  | 95  |
| Conclusion .....  | 96  |
| References.....   | 97  |
| Appendix A: Quality Rating Improvement System Assessment Indicators .....     | 117 |

## List of Tables

|   |    |
|---|----|
| <b>Table 1</b> <i>Operationalization of Variables</i> .....   | 68 |
| <b>Table 2</b> <i>Individual Programs Assessed by Year</i> .....  | 76 |
| <b>Table 3</b> <i>Published Ratings by Number and Percentage of Onsite Observations</i> .....   | 77 |
| <b>Table 4</b> <i>No-Coaching Programs That Received a Program Quality Assessment or Not</i>  | 77 |
| <b>Table 5</b> <i>Precoaching Programs That Received a Program Quality Assessment or Not..</i>  | 78 |
| <b>Table 6</b> <i>Frequency of Precoaching Programs and Their Published Rating Without a<br/>Program Quality Assessment Score</i> ..... | 79 |
| <b>Table 7</b> <i>Research Question 1, Descriptive Statistics</i> .....   | 80 |
| <b>Table 8</b> <i>Analysis of Variance Results</i> .....  | 81 |
| <b>Table 9</b> <i>Tests of Homogeneity of Variances</i> .....   | 84 |
| <b>Table 10</b> <i>Univariate Tests</i> .....   | 85 |
| <b>Table 11</b> <i>Levene's Test of Equality of Error Variances</i> .....   | 87 |

List of Figures

|   |    |
|---|----|
| <b>Figure 1</b> <i>Q-Q Plots</i> .....                        | 83 |
| <b>Figure 2</b> <i>Scatterplot of Change Difference</i> ..... | 86 |

## Chapter 1: Introduction to the Study

Researchers and practitioners agree that high-quality early care and education (ECE) programs significantly influence children's abilities to succeed in school and life (Bartik, 2013; Build Initiative, 2019). However, every state has a patchwork of services rather than a single operator or agency, such as a school district, which provides ECE (Melnick et al., 2017; National Center on Early Childhood Quality Assurance [NCECQA], 2019). Fragmented service delivery makes it challenging for states to assure quality. In response to erratic delivery, Quality Rating Improvement Systems (QRISs) were developed to assess, improve, and communicate the quality level of ECE programs (NCECQA, 2019). See Chapter 2 for a more in-depth explanation of the QRIS. QRISs include multiple components, and each state implements the components in diverse ways. The nature and extent of support and professional development provided to teachers and caregivers vary (NCECQA, 2019). More research is needed to determine the most effective support practices.

The QRIS program for the present study contains four steps: initial rating, quality improvement plan, coaching, and rerating. Each site is assessed with a "Program Quality Rating Score," and then QRIS staff and childcare center programs work together to create a quality improvement plan based on items that scored lower on the assessment. The QRIS then provides up to 3 years of individualized coaching based on the improvement plan. Finally, childcare centers receive a *rerating* assessment to evaluate growth.

Research shows that ECE programs that include intentional coaching score higher on several measures of quality associated with better child outcomes, such as higher

literacy rates, lower perceived negative behaviors, and higher kindergarten entrance exam scores (Bleach, 2015; Forry et al., 2012). Coaching also increases teacher skills and continuously improves the classroom and overall program quality (Melnick et al., 2017; NCECQA, 2019; QRIS National Learning Network, 2013). However, not all early childhood environments receive coaching. Center-based programs, especially those funded by the federal and state governments, typically provide coaching to their early childhood program staff from the inception of the program and on an ongoing basis (Melnick et al., 2017; NCECQA, 2019; QRIS National Learning Network, 2013). Home-based providers and some private programs typically do not receive coaching as a part of their program structure (Melnick et al., 2017; NCECQA, 2019; QRIS National Learning Network, 2013). To raise quality, QRISs provide coaching to all programs and providers upon entrance into the system (Melnick et al., 2017; NCECQA, 2019; QRIS National Learning Network, 2013). One Northcentral state's QRIS provides coaching for all participants.

The findings of the present study may help QRIS staff readjust their interventions to raise quality scores. More knowledge about the differences between groups that receive coaching and those that do not will help researchers and practitioners support early childhood educators and may lead to positive social change. Major sections of Chapter 1 address the study's background, problem statement, purpose, research questions and hypotheses, theoretical framework, nature, definitions, assumptions, scope and delimitations, limitations, and significance. The chapter concludes with a summary.

## **Background of the Study**

Researchers agree that high-quality-rated early childhood programs produce a variety of child outcomes across multiple developmental domains. When children attend a high-quality-rated program, they tend to score higher on measurements of social-emotional development (Landry et al., 2014). Children also demonstrate child leadership skills and strategies, such as independence and cooperation (Forry et al., 2012). Bleach (2015) found that preschool children in high-quality programs scored higher on academic subjects related to school readiness, particularly numeracy skills. Parents, early childhood professionals, and a national agency conducted a 3-year study with 860 children ages birth through 6 years (Bleach, 2015). The results showed that children increased their numeracy skills every year for the full 3 years (Bleach, 2015). Children in high-quality-rated programs also showed reduced stress levels over time, which may have helped prepare them to succeed in school (Vermeer et al., 2010). In addition to child outcomes, many high-quality programs include additional supports for teachers.

Early care and education environments include a patchwork of program types at various levels, including federal, state, local, and community providers (Melnick et al., 2017). Early childhood programs are often implemented and funded through several agencies, such as the Department of Health and Human Services and/or the Federal Department of Education, or through the social services and regular and special education offices at the state level, or through various local and county programs such as local intermediate service districts, and even through community providers such as licensed child care centers, home-based programs, and parochial schools (Melnick et al., 2017).

Early care and education programs typically serve children and families from the prenatal stage through the age of 8 years. Various program guidelines, service options, funding levels, and growth supports lead to differences in program quality scores, making it confusing for policymakers, providers, and families to understand what quality looks like (Baldwin & Wilder, 2014; Bartik, 2013; Melnick et al., 2017).

QRISs were developed to assess, improve, and communicate the quality level of programs serving children from birth through age 8 in all care and education environments (NCECQA, 2019, para. 1). All 50 states have an early childhood QRIS focused on improving program and provider quality (Melnick et al., 2017; QRIS National Learning Network, 2013). When a program serving children from birth through age 8 participates in a QRIS, it improves its overall program quality scores (Abell et al., 2014; Hallam et al., 2017; Hooper & Hallam, 2017; Norris & Guss, 2016; Rentzou, 2017). As a part of QRIS participation, program staff receive coaching that provides feedback on performance and environment factors to increase their skills and continuously improve the classroom and overall program quality (Melnick et al., 2017; NCECQA, 2019; QRIS National Learning Network, 2013). While participation in a QRIS raises a home-based provider's engagement in the QRIS and the early childhood field, there is no guarantee that participation alone fosters continuous improvement (Cortes & Hallam, 2016; Fenech et al., 2010). The recent QRIS research hints at the need for further individualization of a system, rather than streamlining processes (Rentzou, 2017). Faria et al. (2016) also mentioned the lack of research around QRISs and the need for further understanding of how the interventions influenced follow-up rating scores (i.e., any follow-up rating after



the initial rating score). More information about quality, QRISs, and the studies conducted up until now will be explained in Chapter 2.

### **Problem Statement**

Early childhood educators serving children from birth through age 8 seek to continuously improve the quality of the learning environment and curriculum practice (Cortes & Hallam, 2016; Hawkinson et al., 2017). Through the present study, I sought to further the work of Faria et al. (2016) and Yazejian and Iruka (2015). The research problem in the present study was the gap in understanding of the influence of coaching on quality rating scores using a QRIS. Programs with intentional coaching score higher on several measures of quality, leading to better child outcomes; however, not all program types receive coaching before QRIS entry (Bleach, 2015; Forry et al., 2012). All 50 states host a QRIS and provide coaching with variable outcomes (Melnick et al., 2017; QRIS National Learning Network, 2013).

### **Purpose of the Study**

The purpose of this quasi experimental quantitative study was to determine the difference in mean quality scores between two groups of childcare center programs in one Northcentral state: One group received coaching, and one group did not. The independent variable was whether a program or provider received coaching before QRIS entry. The dependent variable was the difference in the mean difference score (initial rating subtracted from the final rating). The next section will cover the research questions in more depth.

## Research Questions and Hypotheses

The archival data contained information indicating that some childcare centers received coaching before the initial rating was done and some did not. The first research question was designed to capture any differences between the two groups (no coaching and pre-coaching) in the initial rating score. The second research question was designed to capture differences in mean change scores after both groups received coaching in the QRIS.

Research Question 1: To what extent is there a difference in the mean initial rating score at QRIS entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry?

H1o: There is no statistically significant difference in mean quality scores between childcare center programs that received coaching before QRIS entry and those that did not.

H1a: There is a statistically significant difference in mean quality scores between childcare center programs that received coaching before QRIS entry and those that did not.

Research Question 2: To what extent is there a difference in the mean difference score (initial rating subtracted from the final rating) at QRIS entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry?

H2o: There is no statistically significant difference in the mean difference score (initial rating subtracted from final rating) at QRIS entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry.

H2a: There is a statistically significant difference in the mean difference score (initial rating subtracted from final rating) at QRIS entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry.

### **Theoretical Framework**

The theoretical framework was Vygotsky's sociocultural theory of learning (Vygotsky, 1978). Vygotsky's sociocultural theory of learning states that people learn more when inside a community of learners than when they are alone (Vygotsky, 1978). Educators have used Vygotsky's theory of social development extensively in the field of education and within early care and education environments to direct teaching and learning activities (Cicconi, 2014). Vygotsky's approach provides detail on how culture and society, specifically through a more knowledgeable other (MKO), facilitate development and learning for children and adults (Liechty et al., 2009; Vygotsky, 1978). Specific practices, such as modeling and scaffolding, are described in Vygotsky's work.

Intentional coaching using MKO approaches positively influence the professional practices of teachers and caregivers in the classroom and home environments (Doyle et

al., 2016; Ronfeldt & Campbell, 2016). The theoretical framework chosen for the present study allowed for insights into how various MKO approaches in the early childhood environment might be useful in raising the quality of a program. The research questions addressed the variations between levels of program and provider scaffolding, which aligned with Vygotsky's approach. Chapter 2 contains a more detailed discussion of the theoretical framework and related research.

### **Nature of the Study**

This study used a quasi experimental pre and posttest design to determine the difference in mean quality scores between two groups of childcare center programs in Northcentral state: One group received coaching, and one group did not (Campbell & Stanley, 1963, p. 6). A quantitative focus is consistent with understanding the relationship of coaching on quality scores within a QRIS (Campbell & Stanley, 1963; Faria et al., 2016; Yazejian & Iruka, 2015). A quantitative focus enabled the exploration of differences, if any, between programs that received *early coaching* (before QRIS entry) and those that did not receive early coaching. The independent variable was whether programs received coaching before QRIS or not. The dependent variable was the difference in the mean difference score (initial rating subtracted from the final rating).

The QRIS participants were determined by self-selection, and participation was voluntary. Participants included staff in center- or home-based early care and education environments around one Northcentral state. I used archival data from the state department of education for programs and providers that already received a quality rating score and a rerating score. I gathered the archival data, checked for complete data for all

childcare centers, cleaned the data, and divided them into two groups: those who received coaching before QRIS entry and those who did not. There was one independent variable and two dependent variables. A one-way analysis of variance (ANOVA) and analysis of covariance (ANCOVA) were used to understand the difference between mean scores of two groups: childcare centers with coaching versus childcare centers without coaching (Green & Salkind, 2014). For more detail on the data collection and analysis, see Chapter 3.

Types of data included the QRIS submeasures, program type, and two rating scores. Sources for data included the state QRIS staff and the state department of education, as well as the QRIS web-based system. The QRIS in the present study used a self-assessment tool combined with the Program Quality Assessment (PQA) tool to provide a final quality rating (Build Initiative, 2019; Great Start to Quality, 2016). The scoring overview of the indicators on the self-assessment tool is in Appendix A. The PQA tool, published by the High Scope Educational Research Foundation, measures quality in four areas: physical environment, daily routine, adult-child interactions, and curriculum and assessment (High Scope Educational Research Foundation, 2003). The data set included initial and follow-up self-assessment scores and PQA scores for each program.

### **Operational Definitions**

The following brief terms are provided for the reader to have a better understanding of the terminology used in the present study. Chapter 3 includes a more detailed analysis of items for coding.

*Coaching*: Onsite, individualized assistance; may include observation, reflection, and feedback, as well as a sharing of knowledge with the teacher or caregiver (Isner et al., 2011).

*Early coaching*: Coaching before QRIS entry; process same as the above definition (Isner et al., 2011).

*Global quality*: A combination of static and dynamic measurements in an early childhood environment (i.e., structural and process quality combined; Connors & Morris, 2015; QRIS National Learning Network, 2013; Rentzou, 2017).

*More knowledgeable other (MKO)*: Someone with more understanding or a higher ability level concerning a task, process, or concept (Liechty et al., 2009).

*Program Quality Rating Score*: The number representing the overall quality rating score within a QRIS. Determined using the three approaches: building blocks, point system, or hybrid (National Center on Child Care Quality Improvement [NCCCQI], 2015). Commonly called the “Star Rating” by those in the system involved in this study.

*Process quality*: Indicators of measurement in an early childhood environment that may change daily, such as adult-child interactions (Connors & Morris, 2015; Rentzou, 2017).

*Quality Rating Improvement System (QRIS)*: “A systematic approach to assess, improve, and communicate the level of quality in early care and education programs” (NCCCQI, 2015, p. 1).

*Rerating score*: The second Program Quality Rating Score used after coaching intervention to show growth in a program or provider’s quality score (NCCCQI, 2015).

*Structural quality*: Indicators of measurement in an early childhood environment that typically stay the same and are determined by stakeholders, such as curriculum and adult-to-child ratios (Connors & Morris, 2015; Rentzou, 2017).

*Technical assistance (TA)*: A mentor practice that provides information on how to access the QRIS and schedule supports, as well as find money and advocacy efforts for quality improvements (Build Initiative, 2019).

*Validation score*: An initial composite score of process and structural indicators as measured by document review; scores indicate whether programs receive an onsite review or not (Great Start to Quality, 2016).

### **Assumptions**

There were several assumptions for the present study. The first assumption was that program directors or owners gave accurate information on the self-assessment screening for their initial rating score. The second was that assessment specialists accurately scored a program during the validation process (measuring the structural quality) and the PQA observation process (measuring the process quality) to determine the final quality score or star rating (measuring the global quality). The third assumption was that the local resource center staff gave accurate information on coaching and technical assistance interventions. The final assumption was that the department of education staff gave the correct data sets for the rating and follow-up rating scores. I received QRIS scores with interval variable data to the hundredth of a point.

### **Scope and Delimitations**

The scope of the present study included center- and home-based programs that received an initial rating, coaching, and then a follow-up rating. The population was intentionally chosen to analyze rating scores and any connection they had with early coaching. Programs without a second rating were excluded. Other populations excluded were those from other states, due to the lack of availability of data. As such, the present study lacks generalizability to other populations or programs. However, it remained a significant study to conduct.

### **Limitations**

There were several limitations to the present study. There was a lack of generalizability, as this was a small sample size of early care and education teachers and providers in one state. Due to the nature of QRIS and voluntary participation, there was a minimal data set for home-based providers. There was no experimental design, control group, or randomization, which would have significantly raised the power of the present study. Only programs participating in the QRIS have quality ratings, limiting the number of programs that I had access to for data analysis. Additionally, only those programs that received an initial rating, coaching, or technical assistance and then a follow-up rating were included.

### **Significance of the Study**

The purpose of the study was to determine the difference in mean quality scores between two groups of childcare center programs in one Northcentral state: One group received coaching, and one group did not. The results may inform future accountability



measures, such as the interventions used within a QRIS and the research tracking teacher, child, and program outcomes. The study may also advise future policy and practice changes, such as supports and best practices used in the classroom, the focus of assessment and evaluation measures, and QRIS processes and supports given to teachers, providers, and caregivers in early childhood. Results may help decision makers determine where to put their financial resources to assure that educators receive the most effective supports and that there is efficient use of finances. The research provides information for programs and providers, coaches, and QRIS implementers on what constitutes effective coaching for change within a QRIS. Additionally, the social change implications may inform future QRIS efforts, advise future policy and practice changes, and help decision makers ensure fiscal responsibility while raising the quality of all early childhood environments. In turn, the local community will also benefit.

### **Summary**

Chapter 1 contained several main points. First, the research is clear that children thrive in high-quality-rated programs. It is not known how coaching helps or hinders quality ratings in early childhood environments in one Northcentral state. Second, the purpose of this study was to determine the difference in mean quality scores between two groups of childcare center programs in one Northcentral state: One group received coaching, and one group did not. I included a description of the research question and hypotheses and explained the connection between Vygotsky's sociocultural theory of learning and the current study. Next, I included a discussion of several factors related to the nature of the study, such as the design and the limitations. The information in the

present study adds to the body of knowledge about program quality interventions in the early childhood environment. Chapter 2 includes research on early childhood program quality, QRISs, and various coaching interventions.

## Chapter 2: Literature Review

Little is known about how coaching influences QRIS quality rating scores. Early childhood educators serving children from ages birth through 8 seek to continuously improve the quality of the learning environment and curriculum practice (Cortes & Hallam, 2016; Hawkinson et al., 2017). Programs with intentional coaching have scored higher on several measures of quality, leading to more significant child outcomes (Banuelos et al., 2019; Bleach, 2015; Forry et al., 2012). However, not all program types received coaching (Bleach, 2015; Forry et al., 2012). All 50 states host a QRIS and provide coaching with variable outcomes (Melnick et al., 2017; QRIS National Learning Network, 2013). The present study was designed to further the work of Faria et al. (2016) and Yazejian and Iruka (2015) to understand how early coaching influences quality ratings within a QRIS.

The purpose of this quasi experimental quantitative study was to determine the difference in mean quality scores between two groups of childcare center programs in one Northcentral state: one group that received coaching and one group that did not. Both groups had received pre and post QRIS ratings. The independent variable was whether a program received early coaching before QRIS entry or not. The dependent variable was the difference in the mean difference score (initial rating subtracted from the final rating).

For this literature review, I analyzed current themes related to coaching and program quality in early childhood environments. First, I include a discussion of where the research was accessed. Then, the theoretical foundation for the present study is explained, including connections to past, present, and future research. Next, I present an

exhaustive look at the current literature related to coaching interventions and QRISs, as well as methodology, studies related to each variable, and what is known and unknown about the variables. Finally, I present a brief statement of conclusions, along with the connection between the gap in the literature and the current study.

### **Literature Search Strategy**

During the literature search, I found multiple articles based on key terms published within the last 7 years. Essential search strategies included using Google Scholar alerts and monitoring several journals, such as the *Early Childhood Education Journal*, *Early Childhood Research Quarterly*, *Professional Development in Education*, and *American Journal of Educational Research*. Based on professional experience, I chose broad keywords to narrow my literature search with my search in the databases. Keywords used in the search included *program quality*, *coaching*, *mentoring*, *professional development*, *high-quality*, and *Quality Rating Improvement Systems (QRIS)*. The databases that I used included Academic Search Complete, ERIC, Education Search, Research Starters—Education, and Teacher Reference Center. After reviewing the literature, I observed broad topics that were related and intertwined and may explain the interaction of variables within the findings of the present study. The search revealed the seminal work that guided the remainder of the literature review.

### **Theoretical Framework**

The theoretical foundation for the present study was Lev Vygotsky's (1978) sociocultural theory of learning. Vygotsky's work was developed in the early to mid-1900s and was translated into English in 1962. The sociocultural theory of learning is

used to explain how culture and society affect a child's development and learning, through active means and with various people throughout the lifespan. This theory of learning laid the groundwork for constructivism. Vygotsky's approach provides details on how others, termed MKOs, facilitate the development and education of a child or adult (Liechty et al., 2009; Vygotsky, 1978). Vygotsky's theory has been used extensively in the field of education (Cicconi, 2014).

Vygotsky's (1978) sociocultural theory of learning forms the foundation of teaching and learning in early childhood environments, including Vygotsky's concept of the zone of proximal development (ZPD). With an appropriate amount of help from others, a child can learn more and perform at a higher level than they can by themselves alone (Vygotsky, 1978). The lower limit of the ZPD describes the level of proficiency that a child receives while working independently. The upper limit of the ZPD describes the proficiency level when the child receives some degree of support from a peer or adult. The verbal instructions—or *scaffolding*—given to a learner help them build new schema in their brain and, over time, gain independence with the materials and increase their skill level (Vygotsky, 1978). The learner, regardless of their age, can learn more and do more when supported by another person. Vygotsky's theory relates to the current study and research questions in several ways.

The current study was aligned with the assumptions of Vygotsky's seminal work. The sociocultural theory of learning builds on a foundation of interactions among people. The learner actively participates with a coach who directs and organizes learning experiences to help the learner attain proficiency (Liechty et al., 2009; Vygotsky, 1978).

The support staff can be anyone who has more knowledge about the topic of focus, such as a peer or an adult. They use scaffolding to adjust the level of support given, over time, to increase proficiency (Vygotsky, 1978). In this study, I sought to understand the relationship between the support staff, ZPD, and scaffolding within a QRIS program. I sought to uncover insights into how the MKO, in the form of a coach, influences practices within the early childhood workforce.

Program providers increase their program quality through various learning experiences. Early care and education professionals need experts to help them through active learning with others, such as a coach. The initial QRIS rating is the lower limit of the ZPD or the level of proficiency that the provider reaches without help from an outside source. The rerate score signifies the upper limit of the ZPD, or the proficiency level achieved with the scaffolding provided by a coach. The research questions for the present study directly related to and built on Vygotsky's theory:

- RQ1: To what extent is there a difference in the mean initial rating score at QRIS entry between childcare center programs that receive coaching and childcare center programs that do not receive coaching prior to QRIS entry?
- RQ2: To what extent is there a difference in the mean difference score (initial rating subtracted from the final rating) at QRIS entry between childcare center programs that receive coaching and childcare center programs that do not receive coaching prior to QRIS entry?

The research questions referred to when a participant might receive scaffolding. Some programs, because of their type and funding, receive early coaching before QRIS entry. Other programs, mostly home-based programs, do not receive early coaching. The initial QRIS score is the lower limit of the ZPD, and the rerating score is the upper limit of the ZPD. In this study, I focused on adults within a broader learning system rather than with children in a classroom to build on Vygotsky's theory. Results have shown that the QRIS structure helps providers become more independent in raising their overall program quality while decreasing scaffolding supports over time. The next section includes an extensive introduction, review, and synthesis of the variables in this study.

### **Literature Review Related to Key Variables**

First, I explain a brief history of the problem, and then I describe research related to four key constructs of interest. Next, I describe research related to the methodology, and I justify the use of the independent and dependent variables. Finally, I synthesize studies related to variables, including what is known, what is controversial, what remains to be studied, and studies about the research questions.

### **History of the Problem**

Historically, research on the problem under study has progressed through five stages. Descriptive statistics were initially used to explain various QRISs, including programs, processes, and people who participated (Fiene et al., 2015). Then, researchers focused on the number of quality programs within a system (Cortes & Hallam, 2016). More recently, studies have been conducted that were focused on using experimental designs to understand the effects of QRIS participation and interventions on rating scores

(Hawkinson et al., 2017; Hooper & Hallam, 2017). Some studies have focused on measuring changes within a system during implementation. Other researchers have turned to quality improvement efforts and outcomes (Faria et al., 2016).

Researchers studied QRISs at inception using descriptive statistics. Descriptive studies were focused on who participated, when they entered, and the purposes for QRIS creation (Ronfeldt & Campbell, 2016). Researchers found that goals for QRIS creation include providing funds for low socioeconomic status children to attend high-quality programs, accountability, and teacher education assessment (Build Initiative, 2019; QRIS National Learning Network, 2013; Ronfeldt & Campbell, 2016). Descriptive studies also were used to describe QRIS processes, such as standards, supports and infrastructure, monitoring and accountability, financial assistance, and engagement and outreach (Connors & Morris, 2015; Faria et al., 2016; QRIS National Learning Network, 2013; Zellman & Fiene, 2012). Researchers also used descriptive studies to highlight other information in a QRIS, such as environment, interactions, academics, state licensing standards, and curriculum indicators (Compendium of Quality Rating Systems and Evaluations, 2010; Hegland et al., 2011; Zaslow et al., 2011). As the QRIS field grew, researchers studied quality rating scores and thresholds for predicting outcomes.

Researchers and practitioners explored quality thresholds that might predict child outcomes. Some studies focused on the number of quality programs, but no information focused on what quality meant or who participated and why they chose to attend. Most QRISs align with their state standards, but each state has a different scoring system. The score depends on the system that one participates in, not necessarily one's quality rating



score (Build Initiative, 2019; Caronongan et al., 2011; QRIS National Learning Network, 2013). QRIS participating programs scored higher on the Early Childhood Environment Rating Scale (ECERS), Classroom Assessment Scoring System (CLASS), and Early Language and Literacy Classroom Observation (ELLCO) than nonparticipating programs. However, that does not mean that QRIS participation results in higher scores. This may mean that participating programs have higher ratings when they enter the system (Jeon et al., 2014). Many of these studies show differences (or no differences) in child outcomes for differently rated programs (Arteaga et al., 2014; Bleach, 2015; Fenech et al., 2010; Jeon & Buettner, 2015). Setodji and Le (2013) found quality thresholds in infant and toddler environments. When assessors scored environments past a specific limit, there were no more differences in child outcomes. Providers and those who support them may not have to aim for the highest score possible but instead shoot for the quality threshold score, saving participants time and money. As time went on, researchers began to look at using an experimental design with control groups.

Very few QRIS studies have used an experimental design. While an experimental design is more valid, there are ethical issues involved when dealing with children (Campbell & Stanley, 1963). Non QRIS participating programs do not often join research studies. Researchers using an experimental design seek to understand if QRIS interventions change rating scores (Boller et al., 2015). While experimental designs with controls are not new to early childhood, they are novel within a QRIS. Landry et al. (2014) is one example of a common experimental design study in early childhood. Three groups of teachers were formed, all of whom served low socioeconomic status children 2

to 3 years of age. The first group was a control group, the second received a new social-emotional-based curriculum, and the third group received the new curriculum, further social-emotional support, and feedback. Researchers found no difference in cognitive abilities among children, but children in both intervention groups outperformed the control group on social-emotional behaviors; teachers in intervention groups had more predictable schedules, more supportive child feedback, and more cognitively stimulating activities compared to controls (Landry et al., 2014). An experimental design matches classroom or program-level interventions; however, using it in a statewide QRIS is often more difficult. Boller et al. (2015) conducted one of the only studies done on QRIS with control groups. Boller et al. compared the quality rating scores, over 6 months, of both programs that participated in the QRIS and those that did not. The results showed higher observed quality scores for those who participated, but no noticeable impact on the overall QRIS rating score. Researchers also collect data on QRIS creation and changes during implementation.

Researchers have conducted validity studies and program evaluations to understand QRISs and their effects. Researchers using program evaluations have sought to understand the changes in implementation and the impact on the people and processes (Elicker et al., 2013; Schulman et al., 2012; Tout, 2013). Some program evaluations have highlighted changes during implementation as well as quality rating scores (Elicker et al., 2013; Schulman et al., 2012; Tout, 2013). Few QRISs use a validation system where they monitor the quality and fidelity of execution (Fenech et al., 2010; Zellman & Fiene, 2012). Most QRIS staff monitor and evaluate after a predetermined number of years,

when they would like to make changes, or when they must justify their efforts to mandating entities (Fenech et al., 2010). Changes to QRIS processes, support, and scoring systems directly affect data and findings within a QRIS study (Faria et al., 2016). More researchers are being asked to monitor and study these changes. Some QRISs use collaborative evaluation processes to ensure quality and fidelity, with multiple layers of stakeholder involvement (Elicker et al., 2013; Schulman et al., 2012; Tout, 2013). While most QRISs do not evaluate themselves or their processes, even fewer use a collaborative approach with multiple stakeholders' input to inform changes. The latest research is beginning to focus on tracking QRIS's outcomes in a more in-depth manner.

Currently, there are little data to show that QRIS efforts and interventions directly affect the QRIS rating. Researchers are beginning to enter QRIS research. Faria et al. (2015) designed new research on how quality improvement efforts influence rating and rerating scores. Faria et al.'s suggested study may show whether the planned interventions influence rerating scores, or changes in quality over time. Seeking to understand a complex system like a QRIS starts with describing the methods and participants first, and then filters down into understanding the many interventions and their effects. As one can see, regardless of past approaches, there remains a need for further research on how quality interventions influence outcomes and changes in rerating scores. The gap in research lends credibility to researching the problem. The next section includes a discussion on the variables in the present study.

## **Constructs of Interest**

The next section includes information about four major constructs of interest related to the present study. The first section describes research on program quality in an early childhood environment and child outcomes. Then, I describe the connections between coaching and program quality rating scores. Next, I include a discussion of QRISs and how they affect program quality. Then I introduce the Northcentral state's QRIS used in the present study. Finally, I include a section on studies related to the methodology.

### ***Program Quality and Child Outcomes***

The first section includes information about three broad topics related to high-quality programming and influences on child outcomes, such as school readiness, social-emotional development, and societal effects. The term “school readiness” means students are prepared for academic success upon entering kindergarten (Bleach, 2015; Konza & Main, 2015). Social-emotional development includes several broad areas related to behavior, stress, and health that determine a child's ability to interact with their peers and adults, as well as being aware of their own emotions (Hemmeter et al., 2015). Past research focused on the many societal effects that result from attendance in high-quality early learning environments and hints at the possible changes in policy and practice (Bartik, 2013). First, I will discuss school readiness.

High-quality early childhood programming includes academic subjects such as language, literacy, and math. Jeon and Buettner (2015) discovered that children enrolled in the highest rated QRIS programs made significant increases in language, literacy, and

math objectives. Children from the lowest socioeconomic status group made more significant improvements in these areas than those from populations at risk of school failure (Jeon & Buettner, 2015). Researchers discovered that regularly using coaching promoted higher instances of specific language and literacy aspects, such as child listening, speaking, reading, and writing in high-quality preschool environments (Banuelos et al., 2019; Konza & Main, 2015). High-quality interventions, such as 2 years of attendance in preschool, increased child numeracy skills in several areas (Arteaga et al., 2014; Bleach, 2015). Fundus (2015) found that all children in a preschool program for children aged 3 to 5 that explicitly used the Creative Curriculum for Preschoolers made increases in child academic assessment areas regardless of the program quality score that they received. Program attendance alone, without a coach's help, may increase the time that it takes to improve skill levels. While academic objectives are essential, they are not the only factors included in an early childhood program.

Early childhood programs also seek to increase the healthy social-emotional development of children. When classroom teams focused on improving their descriptive praise, children's negative behaviors over time were reduced, leading to more positive relationships with their peers and adults (Hemmeter et al., 2015). These effects continued at home with adults in a family environment. One study noted that child attendance in a high-quality program reduced children's stress, as measured by the lowering of the levels of cortisol present in their bodies (Vermeer et al., 2010). Reduced stress led to improved health outcomes and encouraged children's growth in academic skills, as well as to build more positive relationships with their peers and adults (Vermeer et al., 2010). Some

measurement tools within early learning environments focus on predicting social-emotional indicators, like those mentioned above (Compendium of Quality Rating Systems and Evaluations, 2010; Sylva et al., 2006). While most QRIS's do not use a social-emotional indicator tool, they are gaining significance in the research on children's future success (Compendium of Quality Rating Systems and Evaluations, 2010). The importance of social-emotional development is a significant motivator in the field of early childhood education, for children and families, as well as the larger communities in which they live.

The influence of a high-quality program goes beyond the classroom, and is longitudinal, both for the child and the family unit. One family-studies program worked with children from ages birth through five and their parents within an urban setting (Benzies et al., 2014). Results after the age of seven showed a significant gain in child language and global development, as well as a substantial rise in parent self-esteem, a surge in families using community resources, and a considerable lowering of child maltreatment levels (Benzies et al., 2014). A related longitudinal study by Bartik focused on preschoolers and their families (Bartik, 2013). Bartik revealed less prison time and an increased high school graduation rate among children who attended a high-quality preschool environment for 1 or 2 years. Attendance in 2 years of preschool produced lower levels of grade retention, fewer rates of child abuse and neglect, less criminal behavior of the parent and/or the child later in life, higher education attainment levels, more families obtaining and keeping health insurance, and higher job skills for all family members (Lambert et al., 2015). While these studies suggest that high-quality early

childhood programming positively influences child outcomes, there continue to be mixed results in some areas of research.

Due to the many variables in measurement tools, philosophies, and alignment issues, there were mixed results related to program quality and child outcomes. Some quality indicators measured by various tools differ in child outcomes. Perlman et al. (2004) showed that small teacher-to-child ratios predict program quality and influence children's outcomes significantly. Torquati et al. (2007) study showed that teacher's attainment of a child development associate degree and an increase in their pay increased program quality and child outcomes, not lower ratio sizes. Higher ratio sizes also did not affect children's cortisol levels, suggesting larger classroom sizes do not change child outcomes (Vermeer et al., 2010). Other studies confirmed that raising program quality scores influenced child outcomes in combination with other factors, such as children's temperament (Vitiello et al., 2012). Due to significant variabilities across different state's QRISs, there were mixed results in child outcomes in family-based care centers related to child leadership, motivation, and confidence levels (Forry et al., 2012). QRIS validation studies and studies related to quality programming rarely track child outcomes, and when they do the results show no significant difference in child outcomes (Fundus, 2015; Mendive et al., 2016). The confusion among the research studies will be discussed in depth later. However, of note at this point is that although there were mixed results, many researchers agree that focusing on increasing quality among these programs is a top priority (Cloney et al., 2016; Fox et al., 2019; Hallam & Bargreen, 2013; Hawkinson et

al., 2017; Tonyan et al., 2017). One way to increase program quality would be using high-quality coaching.

### ***Coaching and Program Quality***

Coaching is a widely used approach with uses and definitions that vary as much as the outcomes. First, I focus on the ambiguities surrounding coaching, suggested definitions, and a general understanding of the types of coaching, and their results. Next, I include a description of the extensive controversy among educators and other coaching professionals about the meaning, practices, and purposes of coaching. Then, a discussion on the broad and general types of coaching used in and out of the education field, including coaching outcomes across three areas: child, teacher, and program outcomes. Finally, I describe the types of approaches used (often defined as coaching) and their results. First, I focused on the ambiguities surrounding coaching.

The term ‘coaching’ means many different things depending on the environment and the coach. There is widespread confusion about what coaching means in the education world. There are no streamlined definitions or frameworks for coaching, as some programs use specific structures based on academic subjects and others do not (Cecconi et al., 2014; Hemmeter et al., 2015; Mendive et al., 2016; Stockall, 2014). Additionally, different coaching programs focus on different things such as raising program quality scores, raising teacher skills, child outcomes, as well as specific indicators, such as staff training or continuous improvement/program evaluation (Banuelos et al., 2019; Epstein & Wilhite, 2015; Iloabuchi et al., 2016; Lambert et al., 2015). The decentralization of coaching creates confusion, which calls for clarification.



Some larger organizations have proposed definitions of coaching. A brief synopsis is included in the next paragraph.

Coaching is an individual or group process that uses a combination of teaching, modeling, and questioning by the coach to help the participant further their learning and development. Various groups and organizations define coaching differently. The International Coach Federation (ICF) is a worldwide organization that trains, certifies, and oversees coaching practitioners (International Coaching Federation, 2016). ICF defines coaching as a series of activities related to relationship building, effective communication, learning, presence, listening, asking compelling questions, creating awareness, planning, and accountability (International Coaching Federation, 2017). Coaching allows the participant to reach their own predetermined personal and professional goals (International Coaching Federation, 2017). The National Association for the Education of Young Children (NAEYC) defines coaching as a designed relationship to help the teacher or program raise their skill level, as well as their classroom and program quality scores (National Association for the Education of Young Children, 2009). QRIS staff defines coaching as a consultant, training, or technical assistance role, primarily seeking to raise program quality scores (Build Initiative, 2019). One Northcentral state QRIS defines coaching as technical assistance for those programs under a specific quality threshold, and mentoring practices for those above that threshold, and uses a series of prescribed methods within a predetermined framework geared toward programs raising their quality rating scores (Great Start to Quality, 2016). These varying definitions add to the confusion about what coaching is and how to do it efficiently

within early childhood education. General types of coaching are explored in the next paragraph.

There are five types of generalized coaching: individual, team or group, modeling, technological, and coaching supervision. Individual coaching may be based on personal or professional development and is usually focused on increasing skill or the ability of the teacher or administrator (Bowne et al., 2016; Curby & Brock, 2013; MacFarlane et al., 2015; Ottley et al., 2015). Team or group coaching uses a combination of activities and discussions, once or multiple times, to improve classroom or program-wide outcomes (Baldwin et al., 2015; Baldwin & Wilder, 2014; Gerdes & Jefferson, 2015; Gunning et al., 2016; Jao & McDougall, 2016; Park & Ham, 2016). Some coaching programs rely heavily on modeling new skills for teachers and then using a combination of practice and reflection to improve skills (Hemmeter et al., 2015; Ottley et al., 2015; Stockall, 2014). Other forms of coaching use technology, such as Bug in the Ear modeling and performance feedback, a combination of videos, phone calls, and emails, and even the use of Facebook for portfolios and performance feedback (Gynther, 2016; Hemmeter et al., 2015; Kabilan, 2016; Stockall, 2014; Zan & Donegan-Ritter, 2014). Most coaching programs do not provide supervision for coaches (International Coaching Federation, 2016). Lack of coaching supervision suggests that coaches were supervising others with no oversight in their practice. Researchers report varying results on each type of coaching intervention.

Coaching influences child, teacher, and program outcomes. High-quality programs using coaching produce positive child outcomes in school readiness, literacy

skills, reduction in adverse child behaviors, increased play, and facilitates healthy social-emotional development (Curby & Brock, 2013; Durand et al., 2016; Hemmeter et al., 2015; Konza & Main, 2015; Lambert et al., 2015; Landry et al., 2014; McLeod et al., 2019). Several research studies found seven benefits for teachers, including increases in teacher positive attitudes toward teaching, increased teacher emotional support, an increase in fidelity to program methods, and increased ability to handle job stress, positive relationship building skills with children, higher classroom organization, language facilitation, literacy practices, and improved attention to literacy instruction and vocabulary support (Bowne et al., 2016; Curby & Brock, 2013; Landry et al., 2014; MacFarlane et al., 2015; Ottley et al., 2015) Program-related outcomes included an increase in structural and overall quality among all program types and funding streams and increased conversations and adult-child interactions in the classroom (Bleach, 2015; Bowne et al., 2016; Child Trends, 2013; Durand et al., 2016; Landry et al., 2014; Oke, 2016; Ronfeldt & Campbell, 2016). The research suggests that high-quality coaching is indeed a way to effect positive change in the classroom and beyond. However, researchers demonstrated mixed results when using coaching interventions in the preschool classroom.

Researchers found that specific coaching interventions might be most effective. Some coaching versus no coaching at all in the classroom positively influenced the program's overall quality rating (Powell & Diamond, 2013). Programs with some form of coaching outscored those with no coaching on quality rating scores (Powell & Diamond, 2013). Group coaching, where the group focused on collaborative inquiry approaches to

raising quality over time, raised quality rating scores and kept those scores at the higher rating for a more extended period than other programs (Cortes & Hallam, 2016; Hawkinson et al., 2017; Setodji & Le, 2013). However, in several studies, there were no improvements with a generalized coaching program, which may hint at the need for a more structured and focused program geared towards reaching the quality thresholds identified in the research literature, such as a specific literacy coaching program (Bowne et al., 2016; Setodji & Le, 2013). Oke (2016) revealed that the physical classroom environment influences the quality rating scores more than coaching, which adds to the confusion on coaching and quality scores within the early childhood field. The previous research focused on coaching interventions alone. However, using multiple learning methods may enhance coaching over time.

### ***Quality Rating Improvement System and Program Quality***

A QRIS is a framework for early childhood education programs to address quality improvement issues. The next section includes a description of background information on the history, purposes, contexts, and an overview of specific limitations within these types of systems. QRISs are not new in the medical field but have become popular and necessary in the early education field throughout several mandates. Next, I explain the purposes for developing a statewide QRIS, such as funding, accountability, and raising quality to serve children and families better. Then, I include an analysis of how each state uses a different framework, including variations on the types of quality they choose to measure, the tools they use, and the scoring systems they deem most appropriate. Finally, I discuss the general limitations of some of these QRISs in the next section. To

understand the complex nature of QRISs, we first begin with a description of the historical presence of QRISs.

Early childhood QRISs grew in response to a presidential initiative. In 2009, President Obama introduced a new initiative titled Race to the Top- Early Learning Challenge Grant where states could apply for money to put towards building and implementing a QRIS to support children most at risk (Build Initiative, 2019). All 50 states now host a QRIS with different efforts and results (Build Initiative, 2019; Zellman & Fiene, 2012). No two systems are the same, which honors their challenges and opportunities (Zellman & Fiene, 2012). Most state-funded preschool standards align with their QRIS, yet some score these programs differently than other funded programs (Build Initiative, 2019; QRIS National Learning Network, 2013). Many state QRISs align with their state licensing standards, and some states require programs to participate in the QRIS if they wish to remain in operation (Connors & Morris, 2015). Currently, all statewide Early Childhood QRIS have similar purposes.

Most QRISs focus on raising the quality of the classroom for children with the most risk, while some systems serve dual purposes. The five primary objectives for creating a QRIS include addressing quality standards for programs and practitioners, provide support and infrastructure, monitoring and accountability, ongoing financial assistance, and engagement and outreach (QRIS National Learning Network, 2013). The Race to the Top grant requires QRISs to allocate funds to higher-quality programs that serve children with the most risk factors (Build Initiative, 2019). States also use their system to provide accountability for funds given and encourage programs to raise their

quality beyond what is mandated (Ronfeldt & Campbell, 2016). Some states also use their QRIS to assess teachers and their teacher education preparatory programs (Ronfeldt & Campbell, 2016). The purpose of QRIS creation drives the framework implementers choose.

Each state system chooses what types of quality to measure, which measurement tools to use, and how to compare the ratings. Each planning and implementation team determines to measure one or more of three types of quality: structural, process, or global (Bleach, 2015; Bowne et al., 2016; Ronfeldt & Campbell, 2016). Structural quality refers to mandated or required aspects of programming, such as curriculum, materials and equipment, teacher education requirements, and adult to child ratios. Process quality represents more ambiguous concepts, such as teacher and child interactions, and is often more challenging to measure. Global quality refers to a combination of indicators measuring structural and process quality. After implementers determine which type of quality to measure, they choose a measurement tool. Tools measure the physical environment, adult-child interactions, social-emotional feedback, academic concepts, health-related aspects, or a combination of these ideas (Compendium of Quality Rating Systems and Evaluations, 2010; Hegland et al., 2011; Neumann & Wright, 2010). Once implementers determine how to measure their chosen quality, they choose how to rate or rank programs in the system (Build Initiative, 2019; QRIS National Learning Network, 2013). In a leveled system, programs must reach all the points in a level to go on to the next. In a point system, programs are rated based on the total number of points within all levels. In a hybrid system, programs reach a predetermined number of points in each

block for a total score (Great Start to Quality, 2016; QRIS National Learning Network, 2013). These decisions directly affect how program users interpret quality within each system.

QRIS limitations include alignment issues, measurement tool issues, and mixed results in the research. Research indicates areas for further development of QRIS's, including aligning professional development, tools and child outcomes, defining whether the program is assessing for accountability or continuous improvement, and using measures that are relevant across cultures (Bassok et al., 2019; Cicconi, 2014; Compendium of Quality Rating Systems and Evaluations, 2010; Fler et al., 2018; Hu et al., 2015; Ishimine & Tayler, 2013; Rentzou, 2017; Zaslow et al., 2011). Different tools measure different things, and the methodology drives tool usage. Some measurement tools use the phone to assess quality, while others use a combination of observations, document reviews, and interviews; still, others use technological measures (Bassok et al., 2019; Bryant et al., 2011; Cicconi, 2014; Compendium of Quality Rating Systems and Evaluations, 2010; Esplin et al., 2019; Fler et al., 2018; Rentzou, 2017). Not all programs were scored or rated equally among systems, and sometimes not even within the same statewide system (Build Initiative, 2019; QRIS National Learning Network, 2013). Participation in a QRIS does not ensure child outcomes or rating improvements. Sabol and Pianta (2015) found no difference in child outcomes for different rated programs. Fenech et al. (2010) found that many current mandates do not support the efforts needed for high-quality programming. Few QRISs have a validation system of their own to ensure quality and fidelity (Fox et al., 2019; Zellman & Fiene, 2012). These

limitations provide plentiful opportunities for further research. Next, I will focus on a state QRIS.

### **One Northcentral State's Quality Rating Improvement System**

The next section provides general background on QRISs from a national, regional, and statewide basis. As explained above, all fifty states now have some QRIS in various points of development. A recent study of several Northcentral QRIS's revealed seven states that were similar in individual characteristics, yet with many and varied differences (Faria et al., 2016). A description of the broad categories of a Northcentral QRIS is included in the next section. First, a comprehensive view of the picture of QRISs within the United States.

While each state QRIS has unique characteristics, many have the same purposes and, as such, national organizations have formed to connect and support these diverse systems. Each state QRIS adopts or creates quality standards for programs and practitioners, provides support and infrastructure, continuously monitors for accountability, offers various methods of financial assistance, and includes means for engagement and outreach among partners, families, and the community (Build Initiative, 2019; QRIS National Learning Network, 2013). The QRIS Compendium group (Compendium of Quality Rating Systems and Evaluations, 2010) provides a database of info on all fifty states, including research projects and the variations within each QRIS structure. The Build Initiative is a national group formed to help support QRISs through providing resources, technical assistance, tools, and networking, as well as research to inform policy and practice (Build Initiative, 2019; QRIS National Learning Network,



2013). The National Learning Network, a coalition of QRISs and related organizations, partners with the Build Initiative to provide information, learning opportunities, and technical assistance to states who have or desire to create a QRIS (QRIS National Learning Network, 2013). These organizations support individual QRISs, while also providing the foundation for furthering the future of the QRIS field and the impact on early care and education. The next section focuses on the Northcentral Region.

Recently, a mixed-methods study revealed that the QRISs in seven Northcentral states vary in six key areas. Some systems use a block system, some a point system, and others a hybrid system for scoring, yet all seven have the highest rating of four or five stars (Faria et al., 2016). Implementation times for each state vary from Indiana in 2011 to Ohio in 2016 (Faria et al., 2016). All states applied for the Race to the Top funds (Faria et al., 2016). Some states in the region used the funds, and some did not, but all used the process as a motivator to begin their QRIS (Faria et al., 2016). Between the seven states, they use five different tools to measure quality. Some of these tools are nationally known, such as CLASS or ECERS-R, and others use a state-developed measurement tool (Compendium of Quality Rating Systems and Evaluations, 2010; Faria et al., 2016). Each state had different criteria and requirements for assessing programs of the highest quality and used different approaches for choosing classrooms for evaluation (Faria et al., 2016). The focus of the present study was on a system in the North Central region.

One Northcentral QRIS shares similarities with other state systems while maintaining unique characteristics. This state's team applied for the Race to the Top

funds in 2011, 2012, and finally received funds in 2013 (Faria et al., 2016). This QRIS changed since its inception in 2012, specifically regarding assessment (Faria et al., 2016; Great Start to Quality, 2016; QRIS National Learning Network, 2013). Initially, implementers chose a point system and quickly realized they needed to change to a hybrid system. After implementation, and in 2013, they reduced the number of requirements on the self-assessment to be able to qualify for an observation (Faria et al., 2016). Programs must receive a four or five-star rating, out of five stars, to be eligible for an on-site observation. One class per age group was observed, and raters use the Program Quality Assessment tool or PQA (Faria et al., 2016; Great Start to Quality, 2016; QRIS National Learning Network, 2013). There is an accelerated system for federal and state-funded programs as well as NAEYC accredited programs (Faria et al., 2016; Great Start to Quality, 2016; QRIS National Learning Network, 2013). More specifics of how the system runs, including support approaches and system changes, will be discussed in a later section. Next, a discussion on the research related to methodology.

### **Methodology**

Next, I focused on the research about quantitative studies on early childhood QRISs. First, I describe the several types of quantitative research done in and among QRISs. Then, I include examples of descriptive studies, those focused on validity issues, tracking attendance and child outcomes, and very few using an experimental design with controls. Finally, I address a need for more research on QRISs in the future. First, an examination of quantitative studies related to QRISs.

Quantitative studies on early childhood QRISs range in complexity from simple descriptive studies, measuring tool validity, and tracking child outcomes. Very few studies on QRISs use an experimental design with controls, due to the nature of the work and the new field of research. Fiene et al. (2015) used a descriptive study on a QRIS. They found that QRIS participation alone increased quality rating scores among center-based facilities and family and group homes. However, the duration of time within the system significantly impacted their rating scores (Fiene et al., 2015). Two factors that significantly increased the rating scores were the amount of scholarship money received per qualifying child and at least five hours a month of technical assistance (Fiene et al., 2015). Researchers did not track any additional information on dosage, duration, or specific activities. Other descriptive studies highlighted issues related to learning supports, support strategies that providers felt were most useful, and whether providers participated or not in those learning strategies (Cortes & Hallam, 2016; Hawkinson et al., 2017; Hooper & Hallam, 2017; Tonyan et al., 2017). Hong et al. measured concurrent QRIS validity and child outcome predictions (Hong et al., 2015). Hong et al. found that individual indicator ratings do predict overall quality scores. However, the QRIS score does not necessarily predict child outcomes later in kindergarten. One study focused on child outcomes measured kindergarten literacy rates of preschoolers who attended a QRIS participating state-funded preschool (Sabol & Pianta, 2015). Sabol and Pianta found growth in the preschool year, from fall to spring, but no increase during the kindergarten year, when compared to other nonparticipating children (Sabol & Pianta, 2015). Boller et al. (2015) was one of the few experimental design studies done on

QRISs. Boller et al. examined the quality rating scores of programs who participated and those that did not, over a six-month period (Boller et al., 2015). The results showed higher observed quality scores for those who attended, but no noticeable impact on the overall QRIS rating score. More research revealed that participation alone in a QRIS might raise quality scores, even if not reflected in a higher QRIS rating (Hooper & Hallam, 2017; Norris & Guss, 2016). While the amount of research within and among QRISs has increased dramatically over the last 10 years, there is a need for more complex studies with varying designs and methodologies to improve understanding in this valuable new field. The next section will focus on past research related to the problem in the present study.

### **Justification of Variable Selection**

In the following section, I list the independent and dependent variables and operationalize their inclusion. The independent variable was whether a program received coaching before QRIS entry or not. Research shows that early childhood education programs that receive coaching increase early child literacy rates, reduce negative behaviors, and increase kindergarten entrance exam scores (Banuelos et al., 2019; Bleach, 2015; Forry et al., 2012). Preschool programs that receive coaching score higher on teacher skill tests and have higher global quality scores (Melnick et al., 2017; NCECQA, 2019; QRIS National Learning Network, 2013). More recently, Banuelos et al. (2019) found that up to 74% of teachers in a coaching program reached their growth goals for language, behavior management, overall classroom management, and instructional strategies (Banuelos et al., 2019). However, not all early childhood

environments receive coaching. It is important to understand the difference between the quality scores of programs that receive coaching and those that do not to understand the true impact of coaching on quality scores.

The dependent variable was the difference in the mean difference score (initial rating subtracted from the final rating). Quality in an early childhood environment is measured in one of three ways: structural, process, or global quality (Bleach, 2015; Bowne et al., 2016; Ronfeldt & Campbell, 2016). Structural quality refers to items that are standard and may be mandated, such as the curriculum used, the equipment and environment set up, and teacher-to-child ratios (Bleach, 2015; Bowne et al., 2016; Ronfeldt & Campbell, 2016). Process quality includes concepts that are harder to assess and may be more fluid, such as adult and child interactions (Bleach, 2015; Bowne et al., 2016; Ronfeldt & Campbell, 2016). Global quality is measured using a combination of process and structural indicators (Bleach, 2015; Bowne et al., 2016; Ronfeldt & Campbell, 2016). Quality measurement tools may focus on one type of indicator in depth or cover a broad range of indicators in a more comprehensive manner. Specific indicators covered in early childhood quality measurement tools include the environment, daily routine, interactions, social and emotional, academic, or curriculum items (Compendium of Quality Rating Systems and Evaluations, 2010). The quality score is dependent on the tool used to measure quality (Hegland et al., 2011). A Northcentral state's QRIS was used to measure quality scores in this study. To date, there are no studies done on the difference in the mean difference scores in a QRIS. The next section provides a synthesis of the variables, including past and current research related to each variable.

### *Synthesize Studies About Variables*

The next section includes a robust synthesis of the studies about the variables. The research on the independent variable, coaching before QRIS entry, was synthesized into four main sections: coaching and outcomes, duration of standard coaching programs, combination coaching programs, and what the research has to say about duration versus process. The dependent variable, the difference in the mean difference score (initial rating subtracted from the final rating), is described and then extensive research is shared on how quality is measured and specifically how the QRIS in the present study chose to measure quality. Next, I include a section on what was known about the variables and what was considered controversial. The final few paragraphs discuss what remains to be studied and the studies about the research questions.

We know that participation in Professional Learning Communities (PLCs) and high-quality coaching experiences increase overall quality scores and child outcomes (Bleach, 2015; Bowne et al., 2016; Child Trends, 2013; Durand et al., 2016; Gerdes & Jefferson, 2015; Landry et al., 2014; Oke, 2016; Ronfeldt & Campbell). However, private programs often lack the funding required to overcome limitations to participation, such as transportation issues in rural areas, lack of program awareness, and lack of services offered to a private childcare program (Gerdes & Jefferson, 2015). We also know that high-quality coaching experiences influence quality in all program types and funding streams. However, there were varying results as to how long a provider or program must participate in these coaching experiences. Some researchers found that regardless of the time spent in a coaching program, some coaching was better than no coaching when it

comes to raising quality in the classroom (Neumann & Wright, 2010). Some coaching programs last as little as a few months, such as the 8-week job shadow and coaching program laid out in Ferguson's (2016) study. Neumann and Wright (2010) described a 6-month coaching and PD program that increased all environmental rating scores compared to programs that received PD only. Other coaching programs last for a year or more and have mixed results.

Most coaching programs studied by researchers lasted about 1 year in length. Stockall (2014) explained an individual coaching program that also used side by side PD for preschool teachers and paraprofessionals in the same classroom. The results of the 1-year program included significant increases in the paraprofessional's understanding of their roles in the classroom and their abilities to support children with negative behavior (Stockall, 2014). Another 1-year program included a Bug in Ear program (Ottley et al., 2015). Participants were given a device they put in their ear and a coach observed them through one-way glass. The coach instructed them on how to deal with children with negative behaviors. The results showed a significant improvement in teacher-child interaction scores (Ottley et al., 2015). However, one study found that 1 year of a coaching program produced no change in quality or child outcomes (Gunning et al., 2016). The first year produced no change, due to lack of time to collaborate, lack of curriculum knowledge, and no equivalent strategies from curriculum to practice (Gunning et al., 2016). In the second year, participants received more coaching, mentoring, and PD. The results showed an increase in student achievement, especially in

the areas of language development (Gunning et al., 2016). Longer coaching programs may produce greater results.

Researchers conducted several multi-year coaching programs with various results. A 2-year system of various coaching in a school-age facility raised all measures of global quality (Baldwin et al., 2015). A more intensive 2-year program, including PD and coaching for teachers and parents, showed increased receptive language, parent esteem, use of community resources, and a decrease in child maltreatment (Benzies et al., 2014). The results lasted up to 7 years after participants matriculated (Benzies et al., 2014). One 2-year program included teacher observations, child assessments, a new curriculum, weekly literacy coaching, monthly in-service, and summer training (Mohler et al., 2009). Results showed an increase in child word knowledge from the beginning of year 1 to end of year 2 and higher kindergarten entrance scores, compared to those without the 2-year preschool program (Mohler et al., 2009). Cortes and Hallam (2016) described a 2-year coaching program that raised all measures of quality and sustained those scores even 3 years after the program ended, compared with groups who did not receive coaching. Another 3-year program included intentional coaching, with aligned PD, and the use of a highly trained consultant (Labone & Long, 2016). Results showed increased global quality scores and more sustained scores over time with longer participation in the coaching program (Labone & Long, 2016). The evidence may suggest that the combination of intentional coaching programs, along with a sustained time investment may influence quality scores.



Coaching alone does not raise and sustain quality scores over time. Some researchers suggest using a combination of methods to accurately assess, support, and encourage quality improvements and meet Licensing and QRIS mandates (Reinking, 2015). Various combinations of supports have been used in the past. A combination of coaching and college courses showed an increase in child literacy scores with a special education population (Shidler, 2009). Teacher observation, child assessments, coaching, in-service PD, and summer training increased preschool children's language and literacy skills and prepared them for kindergarten (Mohler, et al., 2009). Targeted PD and in-class interventions were used in a class with students on the autism spectrum and showed some improvement in scores (Wilson et al., 2012). However, these results were compared to a combination of PD, interventions, and coaching, and scores for teacher collaborative practices, valuing play-based learning, and child behavior significantly improved (Wilson et al., 2012). Another study showed that the combination of lectures, group work, and using a Plan, Do, Study, Act (PDSA) methodology, along with coaching, increased teacher confidence, overall mentoring and collaboration, and decreased barriers to quality improvement (Maski et al., 2014). These studies suggest coaching alone may be less effective than combining coaching with other supports.

Newer research shows us that there may be a hierarchy for the effectiveness of these coaching interventions. One 4-year study tracked several types of supports and their various outcomes (Akiba & Liang, 2016). They found that informal learning with peer coaching in the educational setting increased student achievement scores more than attending conferences. However, attending conferences increased student achievement

scores more than teacher collaboration projects (Akiba & Liang, 2016). They also found that stand-alone PD sessions, college courses, and individual learning made no difference in student achievement scores (Akiba & Liang, 2016). Sheridan et al., proposed a hierarchy of supports. Coaching and mentoring practices do more than in-service PD sessions to significantly raise and sustain quality scores over time, while in-service PD sessions significantly influence scores more than higher education courses (Sheridan et al., 2009). While these researchers may disagree on which support belong in which placement on the hierarchy, they agree that some supports help more than others.

Other researchers in the field of educational coaching and mentoring highlight the importance of an intentional process of coaching, rather than the duration or hierarchy. Generalized coaching programs with no formal structure and shorter time frames did not raise quality scores in any area and were not appreciated by participating classroom staff (Bowne et al., 2016; McLeod et al., 2019; Setodji & Le, 2013). One multi-year program showed results in child literacy within the first year and declines in years 2 and 3 (Shidler, 2009). Upon closer investigation, the cause was described as a dropping from intentional coaching in years two and three. Teachers reported coaching was not as helpful when it was “off-topic” (Shidler, 2009). One study found that effective and highly trained coaches and mentors make a more significant difference with new teacher effectiveness in all areas measured (Morrissey & Nolan, 2015). Teras (2016) cited the need for more intentionality in educational coaching programs and for more structured facilitation from trained facilitators to increase and sustain quality scores. Participants in these programs described several problems and gave solutions for future work. Teachers

believe supports must be intentional, focused, and given by an experienced person, and that the system should switch the focus from the number of hours of PD attendance to the quality of PD and coaching being offered (Gilbert & Harte, 2013; Linder et al., 2016). Researchers and participants agree that intentional coaching from trained facilitators will cause lasting effects for children. Coaching duration and a combination of focused and intentional supports may indeed affect quality rating scores.

Several studies show conflicting results when examining global measures of quality. Hallam and Bargreen (2013) found that coaching, targeted PD, and teacher orientation raised overall scores, while Fiene et al., (2015) showed that scholarship money, incentives, and technical assistance raised global quality scores. One study even showed that quality scores do not necessarily correlate with QRIS scores. Forry et al. measured quality scores in family-based home programs in three different states Forry et al., 2012). The results showed that providers who scored low on instructional supports and environment indicators also had low QRIS ratings, and those with lower instructional supports and a higher score on the environment had a medium rating in the QRIS. Programs who had more teaching experience and higher teacher degree rating, and a higher percentage of participation in professional organizations, tended to score higher on the QRIS rating, even with lower scores on instructional supports and environment (Forry et al., 2012). Structural quality indicators, such as professional organization membership and teacher degree may influence the initial star rating more than process measures, such as adult-child interactions. When center directors were asked what they would prefer measured they said global measures of quality not just structural (Schulman et al., 2012).

Center directors want everything measured, not just the items you can measure on paper or through an interview. Another study showed the measurements of adult-child interactions and QRIS participation predicted scores of the next QRIS rating more than non QRIS participation (Jeon et al., 2014). Childcare center programs and providers who participate in a QRIS system have an edge on quality ratings and resulting incentives than those who do not participate in a QRIS. The wide variability in scores and measures adds to the confusion about what a QRIS rating means. Furthering the confusion, some researchers show variability in quality scores and QRIS ratings even among classrooms within the same site.

Researchers highlighted the wide variability in quality rating scores among the same site, as well as among different QRISs. One study reviewed and compared quality scores in one center at a time (Karoly et al., 2013). They found a wide variance in scores throughout the center. A 27% variance existed in the quality scores among classrooms within the same site and the same age group and a 40% difference among the same site and different age groups (Karoly et al., 2013). They suggested more classroom observations to receive a more accurate QRIS rating (Karoly et al., 2013). However, when center directors were asked how many classrooms they want observed to get to their global QRIS scores, they reported they wanted as few as possible (Schulman et al., 2012). Some QRISs limit the number of classrooms they assess due to funding and others due to the guidance of researchers on the use of thresholds to predict quality (Build Initiative, 2019). Another study showed the variability among global measures of quality across various states with a QRIS. Connors and Morris (2015) cited a wide difference in

quality measures used among states, but in alignment with most Licensing rules. They also found that QRISs included more process quality measures (e.g., Adult-child interactions) than did Licensing rules (Connors & Morris, 2015). All state QRISs reviewed in the present study aligned with their state preschool guidelines (Connors & Morris, 2015). These results suggest a wide variance in the process indicators among states and the advantage that center-based preschools have when being QRIS rated.

The dependent variable was the difference in the mean difference score (initial rating subtracted from the final rating). As stated above, quality was measured through structural, process, or global indicators and depends on the measurement tool used (Bleach, 2015; Hegland et al., 2011; Ronfeldt & Campbell, 2016). One Northcentral QRIS uses a global quality measurement. The initial round of assessment includes structural indicators, such as teacher to child ratios, curriculum, and teacher education and training (Great Start to Quality, 2016; QRIS National Learning Network, 2013). The second round of assessments includes process quality measures, such as environmental factors, daily routines, and teacher and child interactions (Great Start to Quality, 2016; QRIS National Learning Network, 2013). The final score reflects the global quality, or a combination of process and structural measures (Great Start to Quality, 2016; QRIS National Learning Network, 2013). Next, I will discuss how quality scores were assigned within one Northcentral QRIS.

The validation score in the QRIS used in the present study represents a document review of established measures of structural quality (Great Start to Quality, 2016). The initial validation score in QRIS for the present study measures structural quality

indicators. The initial validation score was an original composite score of indicators measured at the time of entry into the system (Great Start to Quality, 2016). Structural quality indicators typically remain constant once an organization adopts them and were easier to assess with document review and telephone conversations (Connors & Morris, 2015; Rentzou, 2017).

There was, and still is, a wide disagreement in which indicators raise quality scores, as observed by the following research studies. Perlman et al. found adult-child ratios and teacher credentials raised quality scores, while Torquati et al. showed that only teacher degrees are effective (Perlman et al., 2004; Torquati et al., 2007). Neumann and Wright (2010) said intentional coaching programs predict quality scores, while Oke said the environmental rating score predicted quality, not coaching (Neumann & Wright, 2010; Oke, 2016). Hallam and Bargreen (2013) described a system of intentional PD, teacher orientation, and a minimum of a Child Development Associate (CDA) degree predicts quality scores. A medical study on stress levels of children in a preschool classroom showed that group size and adult-to-child ratios common structural quality indicators, influenced children's stress levels (Vermeer et al., 2010). While Hong et al. reported that individual structural quality indicators predict quality scores, but they cannot predict child outcomes (Hong et al., 2015). It was not clear which indicators predict quality scores. Adding to all the confusion, one study reported that orientation, help with submitting paperwork, a mixture of coaching and technical assistance, and group PD significantly influenced a provider or program's quality rating score when initially entering the system (Smith et al., 2010). The support a program receives before

their initial validation occurs could influence their initial validation score and later their initial star rating more than the measured indicators of quality.

The star rating was a composite score of process and structural indicators, or global quality indicators, as determined by the QRIS (Connors & Morris, 2015; Great Start to Quality, 2016; Rentzou, 2017). The number representing the overall quality rating score was determined by one of three different approaches: building blocks, point system, or hybrid (NCCCQI, 2015). The final score is commonly called the “Star Rating” in one Northcentral QRIS (Great Start to Quality, 2016). Each QRIS determines which measures of global quality and which scoring approach they will use, which causes variations in results (Great Start to Quality, 2016; NCCCQI, 2015). The research is still unclear as to which measures of quality predict quality scores.

The initial round of assessment, or validation score, includes structural indicators, such as teacher to child ratios, curriculum, and teacher education and training (Great Start to Quality, 2016; QRIS National Learning Network, 2013). The second round of assessment includes process quality measures, such as environmental factors, daily routines, and teacher and child interactions (Great Start to Quality, 2016; QRIS National Learning Network, 2013). A list of all quality indicators measured during both rounds is in Appendix A. The final score, or star rating, reflects the global quality, or a combination of process and structural measures (Great Start to Quality, 2016; QRIS National Learning Network, 2013). Next, I will discuss how quality scores were assigned within one Northcentral QRIS.

Programs enter a QRIS and experience several rating cycles that, hopefully, result in changes in their quality scores. Upon entrance into the online system, a program was assigned an initial rating score (Great Start to Quality, 2016; QRIS National Learning Network, 2013). The re-rating score was determined 3 years after the initial rating (Great Start to Quality, 2016; QRIS National Learning Network, 2013). The Great Start to Quality system has been around long enough for some programs to go through the re-rating process once (Great Start to Quality, 2016; QRIS National Learning Network, 2013). As such, there was little data about re-rating scores in general and none about how the independent variable influences the dependent variable.

Further investigation into the national landscape of research on the dependent variable also provides a gap in understanding, as no researchers studied the current problem (Gerry Cobb, personal communication, July 5, 2017). However, one small study conducted in Miami, Florida revealed the average participation time for programs in a QRIS was 5 months, and in that time, they were rated and then re-rated (Yazejian & Iruka, 2015). The researchers found that center-based programs raised their total points and star rating, indicating a rise in their overall quality score. However, this was one small study that did not collect information about the interventions used and cannot be generalized due to a low sample size. The lack of data within the United States and the system used in the present study for re-rating scores remains a problem to be further studied.

One Northcentral QRIS used a mixture of coaching, yet current gaps remain in tracking how they influenced quality rating scores. Technical assistance was given to



providers, upon request, before entry into the system and if they scored below a certain quality threshold (Great Start to Quality, 2016; QRIS National Learning Network, 2013). A half-day orientation session was offered to all providers before they signed up to help them understand the system (Great Start to Quality, 2016; QRIS National Learning Network, 2013). Providers who initially scored a zero, one, or two-star rating could request further assistance on how to increase scores for all indicators (Great Start to Quality, 2016; QRIS National Learning Network, 2013). Multiple forms of assistance, provided by phone or email, were completed before professional development sessions in person. Providers who scored a three, four, or five-star on their initial rating could request coaching or mentoring services (Great Start to Quality, 2016; QRIS National Learning Network, 2013). Coaching could be done by phone or in person, or sometimes during professional development sessions, such as in PLCs (Great Start to Quality, 2016; QRIS National Learning Network, 2013). The focus of these courses was on raising quality scores and supporting the provider as a learner (Great Start to Quality, 2016; QRIS National Learning Network, 2013). Providers typically stayed on a caseload for 1 year and had the option to renew their work when they re-rated in 3 years (Great Start to Quality, 2016; QRIS National Learning Network, 2013). Currently, QRIS staff track rating scores, touchpoints with providers, and any improvements made, especially when funds go toward those improvements. However, there is no research to date on how these indicators influence quality rating scores.

There is little data on specific QRIS interventions and child outcomes (Bleach, 2015; Faria et al., 2016; Fundus, 2015; Jeon & Buettner, 2015). There are two national

groups associated with tracking and supporting QRIS efforts within the United States (QRIS National Learning Network, 2013). Neither group has any knowledge of research, to date, tracking rating and re-rating scores in the US (Gerry Cobb, personal communication, July 5, 2017). Another gap that remains to be studied is the various coaching and support approaches offered within the different QRISs. There is no unified definition, purpose, or method of coaching within early childhood and that might affect the quality scores and ratings used for research (National Association for the Education of Young Children, 2003; Race to the Top-Early Learning Challenge, 2014). Neither the QRIS National Learning Network, nor the Build Initiative were aware of any research done on rating scores, coaching interventions, and the effects on re-rating scores (QRIS National Learning Network, 2013). More recently, Banuelos et al. (2019) found increases in overall program quality, especially for initially low-rated programs, however, they reported only small improvements for language and behavior management and no other child outcomes, such as academic subjects. The gaps in knowledge require more research, not only for early childhood and QRIS researchers but also for researchers in the field of professional development and coaching.

The next section covers four overarching controversial themes in the field of QRIS research. The first centers on the many purposes for starting a QRIS. The second includes the disagreement as to what constitutes quality in early childhood and whether the quality rating scores within a QRIS reflect that quality. The third controversy relates to who gets what kind of supports, when, and how often within a QRIS. Finally, there are cultural, historical, and national issues when choosing and using measurement tools, as

the scores influence the supports given to those who work with underserved families and children. Each controversy impacted the present study.

The first controversy focused on a discrepancy between intended and perceived QRIS purposes. Initially, the federal government supported the creation and continuation of state-wide QRISs to provide funding for programs that served the lowest socio-economic populations (Build Initiative, 2019; Race to the Top-Early Learning Challenge, 2014). They also desired to increase low socio-economic community attendance in high-quality programs, thereby providing more of a level playing field for children (Build Initiative, 2019; Race to the Top-Early Learning Challenge, 2014). Over time, QRIS initiators came to understand that a QRIS provided the opportunity to raise the quality of all programs, not just those who worked with underserved populations (Build Initiative, 2019; Jeon & Buettner, 2015). Many early childhood staff observed an increase in support for the early childhood profession, such as an increase in quality standards, more financial aid, and overall monitoring (QRIS National Learning Network, 2013). QRIS programs and research added to the knowledge base on early childhood environments and QRISs in general (QRIS National Learning Network, 2013). However, many providers still believe that QRISs were intended to change them into something they were not, catch them at doing something wrong, or take away funding if they choose to be different (Gerdes & Jefferson, 2015; Hooper & Hallam, 2017).

Further complicating these beliefs are external programs that use the ratings within a QRIS for other purposes, such as accountability for their teacher education preparatory programs (Ronfeldt & Campbell, 2016). While there may be multiple and

varied purposes for the creation and implementation of a QRIS, they remain controversial among many populations. The controversy between QRIS purposes may influence the present study unless the goal of the specified QRIS is clear. Another debate in early childhood is the idea of what constitutes quality.

The second controversy was about the debate among early childhood professionals and practitioners about what quality means and how to measure it over time. Some QRISs measure structural quality, such as curriculum, assessments, and teacher to child ratios, while others measure process quality, such as adult and child interactions, while still others measure global quality, which is a mixture of both (Compendium of Quality Rating Systems and Evaluations, 2010; Connors & Morris, 2015; QRIS National Learning Network, 2013; Rentzou, 2017). Some states use measurement tools focused on adult-child interactions, while others focus more on observing the physical environment, and still, others use phone interviews and checklists to assess quality (Compendium of Quality Rating Systems and Evaluations, 2010; QRIS National Learning Network, 2013). There is no standard scoring tool used among states to compare, making the debate about what is quality even more unclear (Compendium of Quality Rating Systems and Evaluations, 2010; QRIS National Learning Network, 2013). QRIS program ratings depend on the scoring system each state adopts, but the score of the quality measurement tool used may be different (Ishimine & Tayler, 2013; Zaslow et al., 2011). Various states may use the same measurement tool, yet their QRIS rating scores might not be the same. Researchers also warn of alignment issues between QRIS indicators and child outcomes that need to be addressed (Fenech et al., 2010; Sabol &

Pianta, 2015). The controversy about what quality means may influence the current study if the definition of quality is not clear. While there is new research in early childhood because of QRIS creation, the definition of quality remains muddy.

The third controversy centered around the debate about who gets what supports and when in a QRIS. QRIS staff support participants at the program level more than the classroom or individual teacher level, thereby limiting the coaching involvement with individual teachers responsible for supporting children (Build Initiative, 2019; QRIS National Learning Network, 2013). These limitations may have an impact on quality. The current tools, although many and varied, do not assess nonteaching staff and office support or other professionals who work with children in and out of the classroom, which directly affects the quality of the environment (Guernsey & Oschorn, 2011). There were also significant gaps measuring family and group home environments and those that support infant and toddler learners (Sandstrom et al., 2011). The controversy around who gets support and who does not may or may not show up in the current study and affect the data. The dosage of support provided relies heavily on a quality rating score that may be inaccurate due to mismeasurement of specific aspects of quality or not measuring inclusive environments.

The fourth and final controversy in the QRIS research related to the efforts made to be culturally appropriate when measuring quality. There are historical, cultural, national, and sometimes international issues to be aware of when choosing and using an early childhood measurement tool that also affects reliability, and therefore support approaches offered (Cloney et al., 2016; Ishimine & Tayler, 2013; Saracho, 2015).

Programs scoring low on QRIS rating tools were offered fewer support approaches, thereby affecting their ability to raise quality over time. Cloney et al. (2016) showed that high-quality environments, using positive coaching approaches, not only significantly influenced their program quality score but also helped increase child outcomes among lower socio-economic populations. However, they also found that these high-quality coaching approaches were less often provided in these high need areas (Cloney et al., 2016). Programs who might score lower, due to cultural issues, may receive less coaching support. Measurement tools do not always remain relevant across diverse cultures, and therefore cannot be used in an accurate way in the whole nation (Cicconi, 2014; Flee et al., 2018; Hu et al., 2015). The controversy around whether QRIS measures are culturally appropriate may show up in the current study in specific populations of the state or among certain provider types.

### ***Studies About the Research Questions***

The following section includes information related to the research questions. First, I briefly describe various research studies related to coaching. While the researchers disagreed on how long participants should take advantage of coaching, they did agree that some coaching was better than no coaching. Center-based environments hosted the bulk of conducted research, and, more recently, a few studies were conducted on home-based programs. Then, I discuss what the researchers presented as alternative predictors of quality in home- and center-based environments.

There were many studies done on coaching duration and what outcomes if any there were at the program, teacher, and child level. Here, I highlighted just a few noteworthy

examples. One quantitative study looked at two groups, an experimental and a control group. One group received training and intervention and the other received training, intervention, and ongoing coaching (Wilson et al., 2012). Only group two showed an increase in all teacher-measured indicators, including more teacher collaborative practices and an increase in valuing play-based learning (Wilson et al., 2012). Another mixed-method study followed twenty-four teachers through seven different coaching sessions over 2 years and found an increase in all child outcomes measured (Mohler et al., 2009). Shidler (2009) studied 360 children in twelve classrooms over a 3-year coaching program. Results showed no positive change in year one. In year 2 they added more coaching and mentoring, which caused a significant increase in child literacy rates (Shidler, 2009). Another qualitative study focused on coaching within a QRIS (Gilbert & Harte, 2013). Participants reported the need for clear roles and expectations, individualized needs in PD offerings, and specific training for coaches. The data also showed no meaningful change in any child outcomes for the first year of the program (Gilbert & Harte, 2013). These studies may suggest a need for intentional and lengthy coaching programs to raise quality scores and optimize child outcomes.

Researchers reviewed literature and found that family or home-based programs have lower education rates, higher years of experience, and less turnover in staff (Susman-Stillman & Banghart, 2011). Additionally, small increases in instructional supports and higher attendance in professional organizations increased quality scores by 12% among family-based programs, especially when they included minimal literacy instruction (Forry et al., 2012). One quantitative study of 89 family homes and 92 centers showed various

predictors of quality (Holloway et al., 2001). They found different predictors of quality for home-based versus center-based programs. Family Child Care predictors of quality included having activity areas, encouraging dramatic play, no worksheets present, and scheduled parent and teacher conferences (Holloway et al., 2001). Center-based predictors of quality included small group size, higher staff education, no videos, less large group time, less teacher-led academic time, attention to fine motor skills, and the number of parent-teacher conferences (Holloway et al., 2001). Historically, there have been many studies done in the center-based programs that have access to coaching, as they typically receive more money to support research. The present study builds on the work of Faria et al. (2016) who suggested more needs to be done around QRIS research.

### **Summary**

Major themes presented in Chapter 2 include program quality and measurement tools, coaching interventions, QRISs, and one Northcentral state's QRIS structure. High-quality environments influenced children's school readiness, language and literacy, and social-emotional issues. Coaching interventions helped raise the overall program quality score. Not all program types and funding types received automatic coaching, however, all QRIS participants do receive those supports. Participation within a QRIS may or may not influence the overall program quality score, yet participation does not negatively influence the score either. There were gaps related to QRIS interventions, specifically coaching, and the effects on re-rating scores (Faria et al., 2016; Sheridan et al., 2009). I explored to what extent coaching affects quality scores within one Northcentral state's QRIS and extends the work of Faria et al. (2016) and Yazejian and Iruka (2015). Chapter



t3 includes a discussion on the research design, rationale, and methodology to cover the gap in the literature.

### Chapter 3: Research Method

Little is known about how coaching affects quality rating scores in one Northcentral QRIS. The purpose of this quasi experimental quantitative study was to determine the difference in mean quality scores between two groups of childcare center programs in one Northcentral state: One group received coaching, and one group did not. Both groups had pre and post QRIS ratings. In this chapter, I discuss several topics related to the research design and methodology. First, I review the research design and rationale, including a description of the variables, the chosen research design, constraints that I faced, and a defense of my design. Then I present an explanation of the methodology, including sampling procedures, a justification for using archival data, and data collection strategies. Next, I include a brief discussion of instruments used in past research and their relevance to the current study, as well as an operational definition for each variable. Then, I describe my data analysis procedures. Next, I offer a brief explanation of the threats to validity, including internal and external validity issues. Finally, I briefly discuss any ethical considerations for the present study.

#### **Research Design and Rationale**

The present study was a quasi experimental, pretest-posttest design using secondary data analysis and one-way ANOVA and ANCOVA. The independent variable was whether a program received early coaching or not before QRIS entry. The dependent variable was the difference in the mean difference score (initial rating subtracted from the final rating). The pretest-posttest design was appropriate for examining a cause-and-

effect relationship between the independent and dependent variables and was effective in answering the research questions (Campbell & Stanley, 1963, p. 7).

Archival data were used, and there were no random assignments in the already existing groups. All programs and providers entered the system voluntarily, and some childcare center programs in this Northcentral state chose not to enter the QRIS process. A design with experimental control groups would not have worked. Inferential statistics were summarized to describe the relationships between the independent and dependent variables. Data options were limited to what was previously collected. Time and resource constraints applied when receiving data from the state department of education.

Most research on QRISs used descriptive studies only, and very few used an experimental design. Descriptive studies on QRISs covered concepts such as the types of supports offered to providers, barriers to support, the supports that best help family providers grow, and the supports that encourage provider engagement in the system and with other professional organizations (Cortes & Hallam, 2016; Fiene et al., 2015; Hawkinson et al., 2017; Hooper & Hallam, 2017; Tonyan et al., 2017). Hong et al. (2015) conducted a study on concurrent validity on a measure used within a QRIS and found that individual indicator scores, measured separately, do predict global quality rating scores (Hong et al., 2015). Very few studies focused on the connections between interventions and outcomes. One study found that child attendance in a QRIS-participating preschool predicted growth in child development measures from the preschool year to kindergarten but not throughout the kindergarten year (Sabol & Pianta, 2015). Two cases to date used an experimental design within a study on a QRIS. Both

found that participation within a QRIS raises quality scores, regardless of how long one is in the system, while one found an increase in quality scores but not in the QRIS total rating score (Boller et al., 2015; Yazejian & Iruka, 2015). The current study, like the studies just referenced, did not use a true experimental design.

### **Methodology**

The next section includes an in-depth look at concepts related to the methodology such as population, sampling and sampling procedures, archival data, instrumentation, and ethical and validity issues.

#### **Population**

The target population included childcare center programs that received an initial rating and a follow-up rating score in one Northcentral state's QRIS. Center-based environments were referred to as "programs," while home-based environments were referred to as "providers." The target population size for center-based programs and home-based providers was 400. The population size was in alignment with a previous study done by Yazejian and Iruka (2015), in which they used a population size of 412 total programs, with 342 centers and 70 family childcare environments. They studied all environments with one or two ratings over 5 years. The different environment types (i.e., home- and center-based) were aggregated for analysis.

#### **Sampling and Sampling Procedures**

I used data from QRIS programs that received an initial rating and follow-up rating score and coaching in between. Information was extracted from archival data collected from another organization. Programs with only one rating were excluded.

G\*Power 3 was used to calculate the sample size and sampling procedures (Faul et al., 2007). I chose the effect size of .5 to observe a significant effect on data that had not been studied thoroughly in past literature. I chose the standard significance level of  $p = .05$  and the standard power level of .8. I used G\*Power 3 to calculate the sample size to be 74 programs (Faul et al., 2007).

### **Archival Data Procedures**

The present study used an archival data set. In this section, I first explain the original study that collected the data before moving on to the present study. Participants in the main study were recruited to be a part of the QRIS through various methods including media releases, direct mail and letters, licensing communications, and contact at professional development events such as conferences (Great Start to Quality, 2016). Most programs voluntarily participated in the system, such as privately owned center- and home-based programs (Great Start to Quality, 2016). However, federal, and state-funded programs, or those with national accreditation, were required to participate and received incentives for doing so (Great Start to Quality, 2016). Next, I will explain the data collection methods for the main study.

There were several data collection steps involved in the initial study. First, all programs created an account in the system and updated their program profile (Great Start to Quality, 2016). Then, administrators took a self-assessment survey and uploaded documents into the online system (Great Start to Quality, 2016). See Appendix A for indicators included in the assessment. Once documents were uploaded and submitted, an assessment specialist reviewed the records to validate the checked indicators (Great Start

to Quality, 2016). The star rating was then automatically generated through the online system. If a program received a rating of 1 or 2, it received a generated report. The program might or might not have received a follow-up visit or phone call to help staff create an improvement plan (Great Start to Quality, 2016). If a program received a 3, 4, or 5-star rating, it was contacted to schedule an observation visit, which lasted 4 to 6 hours. After the visit, another report was generated, and the provider received a follow-up visit in which a coach provided feedback and helped them create an improvement plan (Great Start to Quality, 2016). Eligibility to apply for a follow-up rating occurs every 3 years. The process repeats and results in a rerating score.

Several procedures were followed to gain access to the data set. First, I applied to request data from the state department of education internal review board. At the same time, I applied for Institutional Review Board (IRB) approval through Walden University, according to communication received directly from Walden University's IRB office. Once approval was received from both institutions, I organized and analyzed the data.

### **Instrumentation**

The High Scope Educational Research Foundation created three instruments used to assess quality in the main study. The latest versions were established in 2000. They are the Preschool Program Quality Assessment (PSPQA), Infant Toddler Program Quality Assessment (ITPQA), and the Family Child Care Program Quality Assessment (FCCPQA). All three measures were appropriate for use in the present study as they assess measures of global quality and fit within the predetermined QRIS design. Each

measure used the same 1-5 scoring scale, and the chosen measure depended on the environment type.

Researchers since 1993 have used PQA tools in over 800 classrooms for various research projects (High Scope Educational Research Foundation, 2018). The published research confirms high interrater reliability and internal consistency, as well as an elevated level of alignment with other program, classroom, and child assessment measures (Schweinhart et al., 2005). The PQA tool is well known in the early childhood field. Researchers in the Perry Preschool longitudinal study tracked attendance in preschool programs that used the PQA measure (Schweinhart et al., 2005). They followed children from age 3 or 4 through the present day (Schweinhart et al., 2005). Additionally, the Great Start Readiness Program study started in 1995 and used the PQA tool for program evaluation and growth plans and is still producing positive child outcomes (Schweinhart et al., 2005). Researchers use PQA tools to assess program quality and collect relevant data. The QRIS staff involved in this study chose to use the PQA tools as an observational measure to track quality (Build Initiative, 2019; Compendium of Quality Rating Systems and Evaluations, 2010; Great Start to Quality, 2016). As such, it was a relevant tool to include in the present study.

### **Operationalization of Variables**

The present study had one independent variable and one dependent variable. The independent variable, early coaching before QRIS entry, is categorical and represented by a 1 for early coaching and a 2 for no coaching. The dependent variable, the difference in

the mean difference score (initial rating subtracted from the final rating), is an interval number and was represented by a score of 1.0 to 5.0, to the tenth of a point.

**Table 1**

*Operationalization of Variables*

| Description | Variable   | Type        | Range of scores                   |
|-------------|--|-------------|-----------------------------------|
| IV          | Coaching before QRIS or not  | Categorical | 1, 2                              |
| DV          | Difference in the mean difference scores (initial rating subtracted from the final rating) | Interval    | 1.0 to 5.0 (to the nearest tenth) |

\*See Appendix A for quality indicators assessed in the PQA tools.

### **Data Analysis Plan**

I used SPSS software for analysis. After I reviewed the data, I removed missing data, as appropriate. For example, many programs and providers only received one rating. These values were removed. I preserved the raw data and saved a cleaned data file for each separate step of the analysis. This helped me protect confidentiality as well as not mix information and provided a more accurate analysis.

### **Research Question and Hypotheses**

RQ1: To what extent is there a difference in the mean initial rating score at QRIS entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry?



H1o: There is no statistically significant difference in the mean initial rating score at QRIS entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry.

H1a: There is a statistically significant difference in the mean initial rating score at QRIS entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry.

RQ2: To what extent is there a difference in the mean difference score (initial rating subtracted from final rating) at QRIS entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry?

H2o: There is no statistically significant difference in the mean difference score (initial rating subtracted from final rating) at QRIS entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry.

H2a: There is a statistically significant difference in the mean difference score (initial rating subtracted from final rating) at QRIS entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry.

First, I conducted descriptive analysis, such as frequencies of respondents for each category, the overall mean scores, and the mean scores across all categories. Then, I used one-way ANOVA for Research Question 1 and ANCOVA for Research Question 2 to explore the differences in scores among the two groups: those who received early coaching before QRIS entry and those who did not.

### **Threats to Validity**

There were several threats to external validity. The PQA measure, while proven valid and reliable, was only one measure of quality and did not go as deep into specific quality indicators as other measurements (Campbell & Stanley, 1963, p. 8; Compendium of Quality Rating Systems and Evaluations, 2010; Fox et al., 2019; High Scope Educational Research Foundation, 2003; Schweinhart et al., 2005). There may have been bias related to program selection, as some participating programs entered in a voluntary status and others did not (Campbell & Stanley, 1963, p. 8; Great Start to Quality, 2016). Recruitment challenges related to voluntary and nonvoluntary status were an issue in several different QRISs (Bassok et al., 2019; Fox et al., 2019). The QRIS in the present study used different measurements, designs, and scoring procedures from other states' QRISs, and, as such, the data cannot be generalized for other QRISs (Build Initiative, 2019; Campbell & Stanley, 1963, p. 8). The study used archival data and was dependent on the items collected; there may have been limitations on what could be studied.

There were several threats to internal validity. The nature of the study design depended on archival data collected from others, which may have affected the validity (Campbell & Stanley, 1963, p. 8). Programs were rated and then rerated, typically within

3 years (Great Start to Quality, 2016). However, some family- and group-home providers waited longer to rerate to get a higher score (Great Start to Quality, 2016). As such, the validity of the study may have been affected regarding history and maturation (Campbell & Stanley, 1963, p. 8). There was no way to control for provider motivation and persistence and the effect that these concepts might have had on the follow-up rating score. It may be that providers were highly motivated to make changes and that their motivation, rather than coaching, helped raise the rerate score. Additionally, there was no way to control how the observation process and the related interventions affected one provider or another, which may have affected the internal validity (Campbell & Stanley, 1963, p. 8). The threats to validity may or may not have influenced the results of this study.

### **Ethical Procedures**

There were ethical considerations to address before starting the study. All relevant institutional permissions, including IRB approvals, were obtained before collecting data. The IRB approval number for this study is 11-25-20-0420457. No live human participants were included in the present study, as it used only secondary archival data.

Upon receipt of the data set, I removed invalid data and coded programs before storage. I saved a file of the original and raw data in a secure environment. Then, I created folders for data in progress, cleaned data, and performed data analysis and descriptive statistics to ensure that the data made sense. I always tracked the location of all data. I stored the data in a secure environment, in paper form as well as digitally. I asked the organization supplying the data for permission before sharing any information,

either written or oral. Only I accessed the original uncoded data. Coded data were made available to the dissertation committee members as needed. All data were destroyed as soon as the study was done, and a copy of the receipt of that process was sent to the governing agency for verification. I remain open to formally presenting the information to the governing organization as requested.

### **Summary**

Chapter 3 covered several topics related to research design and methodology. The present quantitative, quasi experimental, pretest-posttest design used a secondary data set. The sample size for the present study was determined by using G\*Power 3. I included an explanation of the PQA measurement tools, as well as their reliability and validity information, along with the relevance of their use in the present study. Next, a brief discussion occurred about possible statistical tests to run in the future. I also included an in-depth explanation of the ethical issues to consider before collecting data, such as permission, storage, and confidentiality. In Chapter 4, I will present the data and the results of this study.

## Chapter 4: Results

The purpose of this quasi experimental quantitative study was to determine the difference in mean quality scores between two groups of childcare center programs in one Northcentral state: One group received coaching, and one group did not. Both groups had pre and post QRIS ratings. The research questions and hypotheses are listed below.

RQ1: To what extent is there a difference in the mean initial rating score at QRIS entry between childcare programs that received coaching and childcare programs that did not receive coaching prior to QRIS entry?

H1o: There is no statistically significant difference in the mean initial rating score at QRIS entry between childcare programs that received coaching and childcare programs that did not receive coaching prior to QRIS entry.

H1a: There is a statistically significant difference in the mean initial rating score at QRIS entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry.

RQ2: To what extent is there a difference in the mean difference score (initial rating subtracted from final rating) at QRIS entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry?

H2o: There is no statistically significant difference in the mean difference score (initial rating subtracted from final rating) at QRIS

entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry.

H2a: There is a statistically significant difference in the mean difference score (initial rating subtracted from final rating) at QRIS entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry.

Chapter 4 includes study information related to data collection and results. First, I describe the timeline for data collection and recruitment of participants. Next, I present a discussion of the participants, including demographic characteristics and why some programs participated in the QRIS, and the present study, and others did not. Then I offer an explanation of why covariates were used for the second research question and not the first. The rest of the chapter contains descriptive statistics, test assumptions, statistical analysis, any ad-hoc tests performed, and a summary of the answers to the research questions.

### **Data Collection**

This section includes four main parts. In the first part, I explain the recruitment of participants and the timeline for data collection. In the second part, I explain the data that were expected, the data that were received, and some general background information about participants. The third part includes the study criteria and demographic characteristics of participants, including rates of published scores versus those scores of

participants in the actual study. Finally, I justify the exclusion of covariates for Research Question 1 and the inclusion for Research Question 2.

I used G\*Power 3 to calculate the sample size to be 74 programs. I used archival data collected from a previous study. Each participating program or provider received an initial rating and a follow-up rating score between 2014 and 2018. There was a total of 207 programs that fit this criterion in the archival data.

There was a slight discrepancy between the expected data and the data that were received. I expected to receive the funding type to show whether the programs and providers received pre-coaching or not and the initial and follow-up rating scores. Thanks to the participating organization, I received much more, including information on the license type, funding types, program types, validation, and PQA scores, published ratings, and whether programs received pre-coaching or not. These items are covered briefly below.

There were 5,561 different programs with published ratings from the year 2014-2018. Published ratings mean that they went through any level of assessment (i.e., paperwork assessment and/or an onsite observation) to reach their published rating. Table 2 shows the total number of programs that had some level of assessment broken down by the year when they were assessed.

**Table 2***Individual Programs Assessed by Year*

| Year | Number of programs |
|------|--------------------|
| 2014 | 2,087              |
| 2015 | 645                |
| 2016 | 1,394              |
| 2017 | 805                |
| 2018 | 730                |

Not all programs were included in the data set for the present study. While all participating programs received a published Star rating, not all programs received an onsite observation, or PQA. Some programs scored a 1, 2, or 3 on their paperwork assessment, disqualifying them from receiving an onsite observation. Other programs received an “auto-approve” status, meaning that, because they were a federal or state-funded program, or a nationally accredited program, they could self-select a 4-star status without an observation. If they wanted to be certified as a 5-star program, they chose the onsite observation. Few selected this option. There were 3.7% of programs that met the criteria for having two PQA scores within this timeframe. The numbers and percentages of programs with at least one onsite observation and those with no observation at all, broken down into the two groups, precoaching, and no coaching, are shown in Table 3.



**Table 3***Published Ratings by Number and Percentage of Onsite Observations*

|             | Number of published star ratings | Programs with at least one onsite observation | Percentage of published ratings with onsite observations | Programs with published rating and no onsite observation | Percentage of published ratings with no onsite observations |
|-------------|----------------------------------|---|--|--|---|
| No coaching | 3,868                            | 320   | 8.3%   | 3,547  | 91.7%   |
| Precoaching | 1,693                            | 132   | 7.8%   | 1,561  | 92.2%   |

Most of the programs with a published rating and no precoaching did not qualify for an onsite observation. If they received a 1, 2, or 3 on their paperwork assessment, they did not get a PQA. There were 3,457, or 91.7%, of the no-coaching programs that did not qualify for a PQA. Only 8.3% of programs were given an onsite observation score. The 320 no-coaching programs received various scores in the range of 3.00 to 5.00. Table 4 shows the totals and percentages of no-coaching programs that received an onsite observation or not between 2014 and 2018.

**Table 4***No-Coaching Programs That Received a Program Quality Assessment or Not*

| Received PQA or not   | Number of no-coaching programs | Percentage of no-coaching programs |
|-----------------------|--------------------------------|------------------------------------|
| No PQA                | 3,547                          | 91.7%                              |
| PQA score: 3, 4, or 5 | 320                            | 8.3%                               |

Table 5 shows the totals and percentages of precoaching programs that received an onsite observation or not between 2014 and 2018. Of the precoaching programs with a

published rating, 92.2% did not receive an onsite observation. Only 7.8% of programs received a PQA.

**Table 5**

*Precoaching Programs That Received a Program Quality Assessment or Not*

| Received PQA or not   | Number of precoaching programs | Percentage of precoaching programs |
|-----------------------|--------------------------------|------------------------------------|
| No PQA                | 1,561                          | 92.2%                              |
| PQA score: 3, 4, or 5 | 132                            | 7.8%                               |

When I looked closer at the paperwork assessment scores and published ratings of the precoaching programs, there were some clues as to why they would not have more onsite observations. There were 1,561 precoaching programs that did not receive a PQA. Below, I show the breakdown of their paperwork score, keeping in mind that these were programs that qualified for the “auto-approve” option because they were federal or state-funded programs or nationally accredited. The one 5-star published rating was impossible, according to the guidelines of the program. This was an outlier, due to improper manual entry. Table 6 shows the number and percentages of precoaching programs that did not receive an onsite observation and their final published rating.

**Table 6***Frequency of Precoaching Programs and Their Published Rating Without a Program**Quality Assessment Score*

| Published rating | Frequency of programs | Percentage |
|------------------|-----------------------|------------|
| 1                | 8                     | .5%        |
| 2                | 20                    | 1.3%       |
| 3                | 140                   | 8.9%       |
| 4                | 1,392                 | 89%        |
| 5                | 1                     | .3%        |

Of the 5,561 programs that entered this QRIS and were assessed between 2014-2018, only 207 received more than one onsite observation. This was in large part due to federal and state-funded precoaching programs refusing to be assessed and taking the auto-approved 4-star rating. The criteria for selection into the data set for the present study included at least two onsite assessment scores. This explains the large discrepancy between the number of published ratings (5,561) during this period and the number of programs included in this data set (207).

The first question addressed the extent of the difference between mean scores for the initial PQA and the follow-up PQA for two groups. This was simple, and no covariate was needed. However, through the second research question, I sought to understand the change between difference scores of the two groups and, as such, it was necessary to hold the first PQA score as a covariate. More is included on this in the next section on results.

### **Results**

This section includes information related to the statistical tests. First, I begin with a review of the two research questions. Then, for each research question, I explain the test that I chose and why. Then, I include an explanation of each of the assumptions and the

assumption test results. Finally, I present a conclusion statement about the significance of each test.

### Research Question 1

Research Question 1 was as follows: To what extent is there a difference in the mean initial rating score at QRIS entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry? Participants were classified into two groups: no coaching ( $n = 127$ ) and precoaching ( $n = 80$ ). Data are presented as mean  $\pm$  standard deviation: precoaching ( $n = 80, 4.49 \pm .41$ ) and no coaching ( $n = 127, 4.31 \pm .43$ ). Table 7 shows descriptive statistics of the two groups.

**Table 7**

*Research Question 1, Descriptive Statistics*

|                |          |        |                   |               | 95% confidence<br>interval for mean |                |         |         |
|----------------|----------|--------|-------------------|---------------|-------------------------------------|----------------|---------|---------|
|                | <i>N</i> | Mean   | Std.<br>deviation | Std.<br>error | Lower<br>bound                      | Upper<br>bound | Minimum | Maximum |
| No<br>coaching | 127      | 4.3073 | .42524            | .03773        | 4.2326                              | 4.3820         | 3.21    | 4.99    |
| Precoaching    | 80       | 4.4923 | .41629            | .04654        | 4.3996                              | 4.5849         | 2.25    | 4.97    |
| Total          | 207      | 4.3788 | .43036            | .02991        | 4.3198                              | 4.4378         | 2.25    | 4.99    |

A one-way ANOVA was conducted to determine whether there was a statistically significant difference in the mean initial PQA scores for two groups with different exposures to coaching.

**Table 8***Analysis of Variance Results*

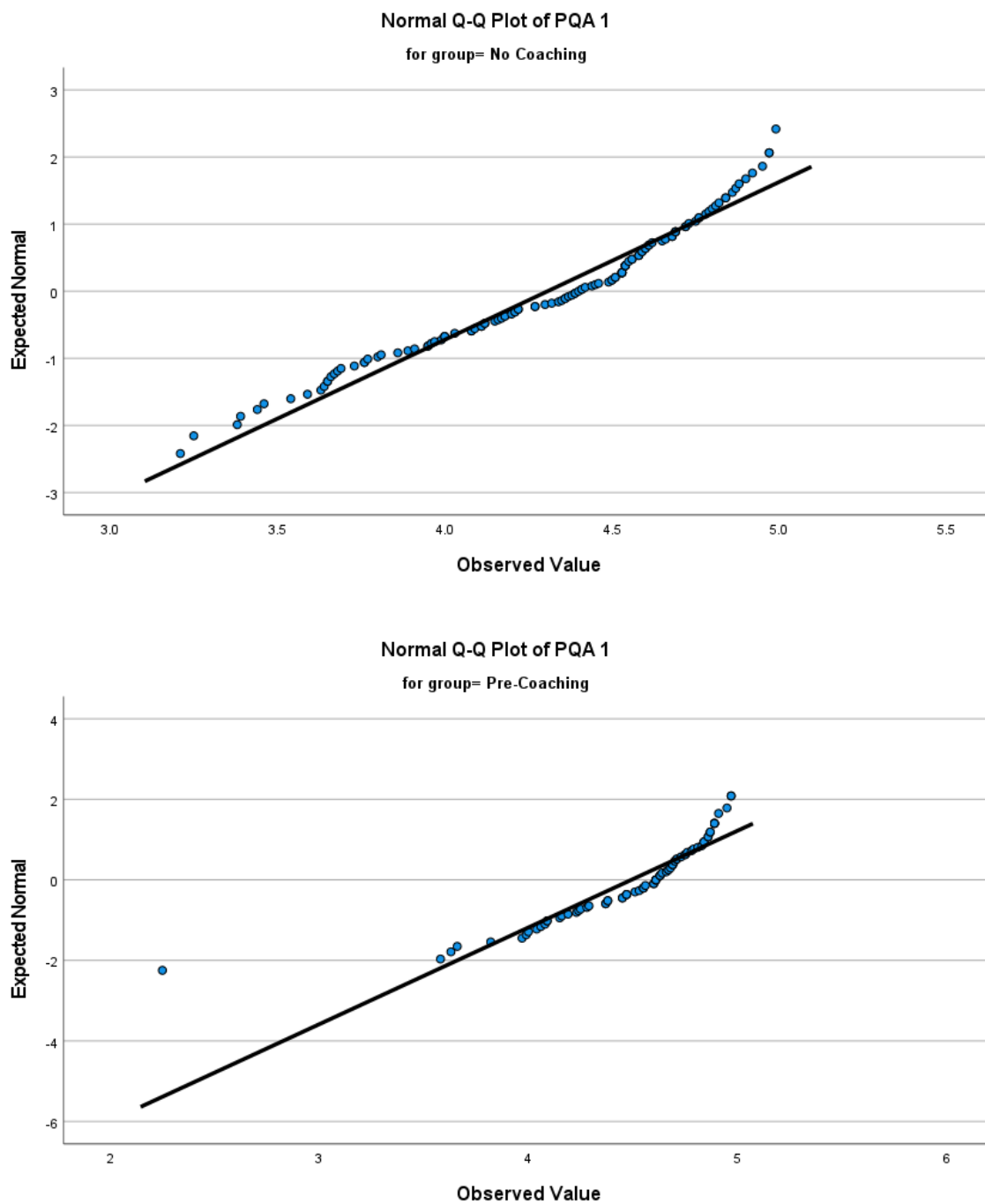
|                |                | PQA 1     |             |          |      |
|----------------|----------------|-----------|-------------|----------|------|
|                | Sum of squares | <i>df</i> | Mean square | <i>F</i> | Sig. |
| Between groups | 1.679          | 1         | 1.679       | 9.434    | .002 |
| Within groups  | 36.475         | 205       | .178        |          |      |
| Total          | 38.153         | 206       |             |          |      |

A general linear model univariate test was run to calculate the effect size and was found not to be statistically significant with a value of partial eta squared = .029. This means that, while the ANOVA results show a statistically significant difference in the mean scores between the two groups, the effect was not large. No post hoc tests were run because there were fewer than three groups to compare. Therefore, I rejected the null hypothesis for the first research question. There was a statistically significant difference in the mean initial rating score at QRIS entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry, even with a small effect size and childcare center programs that did not receive coaching prior to QRIS entry. The alternative hypothesis was accepted because there was a statistically significant difference in the mean initial PQA scores; with the pre-coaching group scoring higher.

Six assumptions needed to be met for the ANOVA test. Each assumption was met. I list them here with a brief description and result.

1. One continuous dependent variable. This included the PQA score; therefore, the assumption was met.

2. Two or more categorically independent variables. The two groups were the precoaching group and the noncoaching group, both independent of each other; this assumption was met.
3. Independence of observations. All program scores were conducted as independent observations in the original study; this assumption was met.
4. No significant outliers. There were two outlier points. The outliers were deemed to be data entry errors and were kept, as transforming them, or deleting them, would have caused significant changes in the results. This assumption was met.
5. The dependent variable was normally distributed for each independent group. The results show an approximated normal distribution based on the q-q plot in Figure 1. This assumption was met.
6. Homogeneity of variances (or the variances in each group were about the same), Levene's test of homogeneity of variances. The high  $p$  value ( $p = .126$ , Table 9) led me to reject the assumptions of variances as significantly different. This assumption was met.

**Figure 1***Q-Q Plots*

**Table 9***Tests of Homogeneity of Variances*

|        |  | Levene statistic | df1 | df2     | Sig. |
|--------|--|------------------|-----|---------|------|
| /PQA 1 | Based on mean                                  | 2.363            | 1   | 205     | .126 |
|        | Based on median                                | 2.568            | 1   | 205     | .111 |
|        | Based on median and with<br>adjusted <i>df</i> | 2.568            | 1   | 194.976 | .111 |
|        | Based on trimmed mean                          | 2.838            | 1   | 205     | .094 |

**Research Question 2**

RQ 2: To what extent is there a difference in the mean difference score (initial rating subtracted from final rating) at QRIS entry between childcare center programs that received coaching and childcare center programs that did not receive coaching prior to QRIS entry? Again, participants were classified into two groups: no coaching ( $n = 127$ ) and pre-coaching ( $n = 80$ ). An ANCOVA was run to determine the effect of coaching interventions on quality rating scores. After adjustment for the covariate (initial rating score), there was a statistically significant difference in the mean change scores (initial rating subtracted from the final rating) between the two coaching groups ( $F = 6.174$ ,  $p = .014$ , partial  $\eta^2 = .029$ ). Table 10 shows the ANCOVA results. Post hoc analysis was performed with a Bonferroni adjustment and the mean difference change was statistically significant.



**Table 10***Univariate Tests*

|          | Sum of squares | <i>df</i> | Mean square | <i>F</i> | Sig. | Partial eta squared |
|----------|----------------|-----------|-------------|----------|------|---------------------|
| Contrast | .676           | 1         | .676        | 6.174    | .014 | .029                |
| Error    | 22.328         | 204       | .109        |          |      |                     |

*Note.* Dependent variable: change difference. The *F* tests the effect of coaching or not.

This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

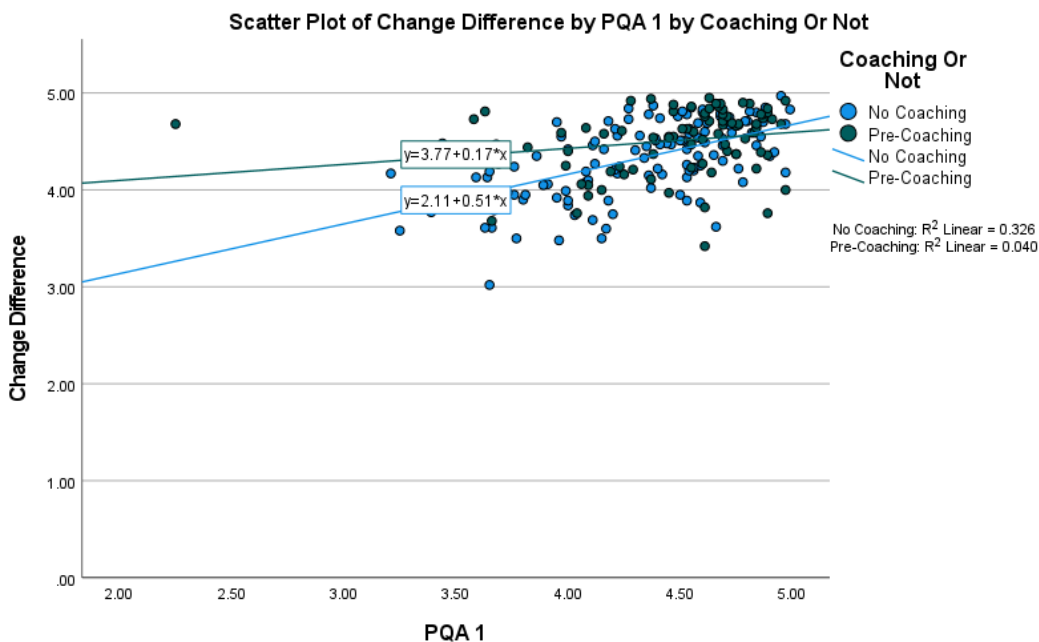
Seven assumptions needed to be met for the ANCOVA test. I list them here with a brief description and result.

1. One continuous dependent variable. This included the PQA score, therefore the assumption was met.
2. Two or more categorically independent variables. The two groups were the precoaching group and the noncoaching group, both independent of each other, this assumption was met.
3. A continuous covariate. The original PQA score was continuous and therefore this assumption was met.
4. Independence of observations. All program scores were conducted as independent observations in the original study, assumption met.
5. There was a linear relationship between the two groups as assessed by visual inspection of a scatterplot. The lines both go up showing a linear relationship. However, the slopes of the lines were not homogenous as seen in Figure 2.

6. There was no homogeneity of regression slopes as the interaction term was statistically significant ( $p = .014$ ). This assumption was not met. There was homoscedasticity and homogeneity of variances as assessed by visual inspection of a scatterplot.
7. The results of Levene’s test led me to fail to reject the null hypothesis of equal error variances: ( $F = .030, p = .862$ ). The data meet the homogeneity of variances assumption.

**Figure 2**

*Scatterplot of Change Difference*



**Table 11***Levene's Test of Equality of Error Variances*

| <i>F</i> | <i>df1</i> | <i>df2</i> | <i>Sig.</i> |
|----------|------------|------------|-------------|
| .030     | 1          | 205        | .862        |

*Note.* Dependent variable: change difference. Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

While the homogeneity of regression assumption was violated, I continued with the ANCOVA as these violations may be a result of the small population size. G\*Power 3 was used to calculate the sample size and sampling procedures (Faul et al., 2007). I chose the effect size of .5 to observe a significant effect on data that has not been studied thoroughly in past literature. I chose the standard significance level of  $p = .05$  and the standard power level of .80. I entered the numbers into G\*Power 3 and calculated the sample size to be 74 programs. While not as much as our expected population of 400 or more, this was significantly less than our total population size of 207.

All assumptions were met except the assumption of slope regressions. Although all assumptions were not met the ANCOVA results were included because the small sample size may have influenced the statistical tests. A larger sample size may have provided a more robust sample to confirm the normal distribution.

### **Summary**

Two tests were run to determine differences between two groups—precoaching and no-coaching—on quality rating scores within a specific Quality Rating Improvement System. The ANOVA test compared the scores for each group on the initial quality rating assessment. Results were that precoaching and the quality rating score were not

independent, which suggests a relationship may exist between pre-coaching and growth in this Northcentral QRIS. The ANOVA showed a statistically significant difference with a range of scores between 4.31 and 4.49, with the pre-coaching group scoring higher. More will be discussed about this significance to providers and programs in Chapter 5.

An ANCOVA was run to determine the effect of coaching interventions on quality rating scores once programs were in the QRIS. After adjustment for the initial rating score, there was a statistically significant difference in the mean change scores between the two coaching groups. This suggests that coaching within this QRIS may be associated with quality rating scores and care for children. Chapter 5 will discuss these results in-depth.

## Chapter 5: Discussion, Conclusions, and Recommendations

The purpose of this quasi experimental quantitative study was to determine the difference in mean quality scores between two groups of childcare center programs in one Northcentral state: One group received coaching, and one group did not. Both groups had a pre and post QRIS rating. The findings of the present study may help QRIS staff readjust their interventions to raise quality scores. More knowledge about the differences between the two groups will help researchers and practitioners support early childhood educators and may lead to positive social change.

The key findings demonstrate that coaching may be associated with the quality rating scores within this QRIS. The ANOVA for Research Question 1 showed a statistically significant difference with a range of scores between 4.31 and 4.49. This means that programs that received coaching prior to QRIS entry scored higher on the initial program quality rating. An ANCOVA was run to determine the effect of coaching interventions on quality rating scores once programs were in the QRIS and had two or more rerating scores. After adjustment for the initial rating score, there was no statistically significant difference in the mean change scores between the two coaching groups. This means that programs that received no precoaching prior to QRIS entry raised their score more significantly than programs that received precoaching. In other words, those programs that previously did not receive coaching raised their quality scores more than those that did receive coaching once they entered the QRIS.

### **Interpretation of Findings to the Empirical Literature**

The present findings confirm several research studies laid out in Chapter 2. Fiene et al. found that participation alone in a QRIS increased a program's quality score (Fiene et al., 2015). The ANCOVA results confirmed that programs that participated in a QRIS did raise their quality score. Several research studies found that some coaching was better than no coaching (Ferguson, 2016; Fiene et al., 2015; Stockall, 2014). The results of the present study indicated that coaching prior to QRIS raised quality scores at the time of QRIS entry compared to no-coaching programs. Programs that received coaching for the first time in a QRIS experienced greater growth rates of quality scores over 2 years. Labone and Long (2016) found that the amount of time that a program participated in a coaching program influenced its quality rating scores and QRIS rating. The ANCOVA results for Research Question 2 show this to be the case: The pre-coaching program had more exposure to coaching practices than the no-coaching program and had higher scores. Once the no-coaching programs received coaching in the QRIS, their ratings grew at a more rapid rate than did the ratings for the pre-coaching programs.

Past researchers concluded that coaching supports would raise the program quality score (Bleach, 2015; Bowne et al., 2016). The supports that they found most helpful were included in this QRIS: technical assistance, financial supports, process coaching, and professional development opportunities (Bleach, 2015; Bowne et al., 2016; Great Start to Quality, 2016). When programs received coaching supports in this Northcentral state QRIS, they raised their quality scores. This aligns with the small study done by Yazejian and Iruka (2015), especially for programs who received lower financial

supports for coaching prior to QRIS entry (Gerdes & Jefferson, 2015). While the findings confirm much of the past research, they do not confirm everything listed in Chapter 2.

Some findings from past research could not be confirmed due to several limitations. For example, past researchers focused heavily on program quality and effects on teacher and child outcomes (Bowne et al., 2016; Hemmeter et al., 2015; McLeod et al., 2019). The scope of the present study included global quality scores of a very narrow criterion of programs within a certain period and did not look at teacher and child outcomes. Much of the past research focused on child outcomes related to school readiness, social-emotional health, behavior issues, and societal outcomes (Benzies et al., 2014; Hemmeter et al., 2015; Jeon & Buettner, 2015). Some researchers found that the more time that a program spent with specific coaching supports, the higher its quality score was (Ferguson, 2016; Fiene et al., 2015; Neumann & Wright, 2010; Stockall, 2014). One study by Benzies et al. showed that participating in coaching programs had lasting changes in quality even 7 years later (Benzies et al., 2014). The present study cannot confirm or disconfirm any past research older than 2 years, as I only looked at programs rated between 2 and 3 years. There were only five programs out of 207 included in the data that were rated over 3 years or more. Many researchers in the past literature highlighted several reasons for higher quality, and coaching was only one factor that they listed (Akiba & Liang, 2016; Boller et al., 2015; Hooper & Hallam, 2017). The present study did not focus on any of the factors listed in these past studies, such as teacher-child ratios, teacher education and experience levels, and diverse types of

coaching supports. As such, I cannot confirm or disconfirm those results. However, I can list several ways that the present study results extend knowledge in multiple fields.

While this was a quasi experimental study, the findings extend the knowledge of QRIS and quality beyond what the profession may have known before. Most QRIS studies have been descriptive in nature (Connors & Morris, 2015; Faria et al., 2016; Ronfeldt & Campbell, 2016). Some studies have focused on quality thresholds to inform financial supports for QRIS developers (Build Initiative, 2019; Setodji & Le, 2013). Numerous studies have looked at validating the systems, processes, assessors, or whole QRIS for effectiveness (Elicker et al., 2013; Fenech et al., 2010; Schulman et al., 2012; Tout, 2013; Zellman & Fiene, 2012). The findings of the present study go beyond descriptive statistics and validation information to inform readers about how coaching influences within the QRIS help programs, specifically those with no coaching supports prior to QRIS entry. Additionally, it extends limited knowledge on how coaching influences home-based and private center-based programs (i.e., no-coaching programs).

### **Interpretation of Findings in Relationship to the Theoretical Framework**

The findings of this study confirm the sociocultural theory of learning laid out by Vygotsky (1978). The sociocultural theory of learning builds on a foundation of interaction between people. The learner actively participates with a coach who directs and organizes learning experiences to help the learner attain the proficiency stage (Liechty et al., 2009; Vygotsky, 1978). The support staff can be anyone who has more knowledge about the topic of focus, such as a peer or coach. MKO supports raise quality



scores. MKOs, or coaches, use scaffolding (coaching) to adjust the level of support given over time, to increase proficiency (Vygotsky, 1978).

Programs that participated in this QRIS raised their quality scores with MKO assistance. The ZPD in the QRIS was a range of scores representing what providers could accomplish on their own without MKO supports (lowest level) and what they could master with MKO supports (highest level). The ANOVA results show that the level of MKO assistance before QRIS entry affected the initial quality rating or lowest level ZPD score. Programs that received the MKO support before entry received higher initial scores than those that did not, meaning that the ZPD scores were significantly different. The follow-up rating, done 2 years post entry, or the proficiency level achieved with MKO support, was different between the two groups. ANCOVA results showed that a program that previously had no MKO support received coaching within the QRIS and raised quality scores (ZPD) faster than those programs that received MKO assistance prior to entry. This means that programs that are now supported through a knowledgeable coach will be able to support children to better outcomes.

### **Limitations of the Study**

The scope of this study included center- and home-based programs that received an initial rating, coaching, and then a follow-up rating. Programs without a second rating were excluded. At 3.7% of the total licensed childcare population, this was a small sample size of early care and education teachers and providers in one Northcentral state. As such, the present study lacks generalizability to other populations or programs. Due to the nature of QRIS and voluntary participation, there was a minimal data set for home-

based providers. There was a quasi experimental design, but no control group or randomization, which would significantly raise the power of the present study. Only programs participating in the QRIS have quality ratings, limiting the number of programs accessible for data analysis.

### **Recommendations**

This section includes a discussion of the recommendations based on the study findings. First, I include recommendations for further study based on the limitations highlighted in this study. Then, I present a brief discussion of the recommendations for methodology changes for future research. The present study used scores from 207 childcare programs. Due to the nature of the QRIS and archival data, there were thousands of programs that were not included in the data set. Hundreds of pre-coaching programs were not included due to the auto-approval system (see Chapter 4 for more information). Because of this small subset of data, I recommend that more PQA data are collected in the future. There were 108 home-based programs included in this data set, making up 2.3% of the total population of licensed facilities of all kinds in this Northcentral state. I would recommend doing another study with a much larger set of home-based and private center-based (i.e., no-coaching group) in the future to see how the data might be different.

There are some methodology implications to consider. The present study included data from programs that voluntarily participated within this QRIS. As such, there was no quasi experimental design. I suggest that researchers conducting future studies try to design, as much as possible, a study with a control group to evaluate further. Control

groups may not be possible, as most QRISs are voluntary; however, should a system require programs to enter in the future, an experimental design might prove more valuable. Due to the voluntary nature of the study population, there were also some issues with recruitment and sampling. Researchers may want to look at data from other QRISs that were not voluntary to see if the findings align. The present study used a short 4-year period to collect data. Future researchers, within this QRIS and others, may want to do a longitudinal study to see if the influences of coaching remain after a longer period.

Other factors may also play a key role in whether programs raised their quality scores or not. I recommend focusing on factors such as teacher education and experience, adult-to-child ratios, and types of curricula used. Future researchers may also want to focus on what factors, other than coaching participation, influence quality rating scores, such as coaching supports and types in various QRISs and coaching within those QRISs. Additionally, this study and past research did not include information about how the QRIS rating affects teacher and child outcomes. This is an area for further research. Finally, this study looked at only one QRIS. Future researchers may want to compare the findings of various QRISs.

### **Implications for Positive Social Change**

The findings of the present study may help the systems, agencies, and organizations that support early childhood staff amend support to all program types, not just the federal- and state-funded programs that already receive coaching supports. This may lead to changes in program quality and better child outcomes. The findings may also inform future QRIS efforts by putting more emphasis on coaching for programs that

previously did not receive it before QRIS entry. Positive social change may also occur by helping decision makers ensure fiscal responsibility while raising the quality of all early childhood environments. In turn, the local community may also benefit.

### **Conclusion**

The purpose of the present study was to determine the difference in mean QRIS scores between two groups of childcare centers in one Northcentral state: One group received coaching, and the other did not. Results showed that programs that received coaching before QRIS entry were rated higher on several aspects of program quality at the time of QRIS entry. Programs that did not receive coaching before QRIS had lower quality scores at the time of QRIS entry. However, they raised their quality scores more than the pre-coaching group, from initial rating to follow-up rating. This means that providing programs with access to coaching that typically have less access to support services helps raise the quality rating scores in this Northcentral state's QRIS. Coaching within this QRIS helped programs with less access to support services prior to QRIS raise their quality scores. Every state in the United States has a QRIS to assess childcare center program quality and provide coaching for teachers. The present study and past research showed that teacher coaching improves program quality scores and improves child outcomes. The findings of the present study may significantly help inform those who support early childhood education and care staff within this QRIS and others. Early childhood educators of all funding types serve children ages birth through 8 and seek to continuously improve their quality, and they deserve all the help that they can get to do that job effectively.

## References

- Abell, E., Arsiwalla, D. D., Putnam, R., & Miller, E. B. (2014). Mentoring and facilitating professional engagement as quality enhancement strategies: An overview and evaluation of the family child care partnerships program. *Child Youth Care Forum, 43*(5), 569-592. <https://doi.org/10.1007/s10566-014-9254-1>
- Akiba, M., & Liang, G. (2016). Effects of teacher professional learning activities on student achievement growth. *Journal of Educational Research, 109*(1), 99-110. <https://doi.org/10.1080/00220671.2014.924470>
- Arteaga, I., Humpage, S., Reynolds, A. J., & Temple, J. A. (2014). One year of preschool or two: Is it important for adult outcomes? *Economics of Education Review, 40*, 221-237. <https://doi.org/10.1016/j.evalprogplan.2011.12.011>
- Baldwin, C. K., Stromwall, K., & Wilder, Q. (2015). Afterschool youth program design and structural quality: Implications for quality improvement. *Child and Youth Services, 36*(3), 226-247. <https://doi.org/10.1080/0145925X.2015.1046592>
- Baldwin, C. K., & Wilder, Q. (2014). Inside quality: Examination of quality improvement processes in afterschool youth programs. *Child & Youth Services, 35*(2), 152-168. <https://doi.org/10.1080/0145935X.2014.924346>
- Banuelos, N. V., Doerfel, M. K., & Stoffel, R. E. (2019). Bright and early: Coaching increases the quality of early childhood programs. *The Learning Professional, 40*(6), 50-53.
- Bartik, T. J. (2013). *Effects of the pre-K program of Kalamazoo County Ready 4's on kindergarten entry test scores: Estimates based on data from the fall of 2011 and*

*the fall of 2012* (Upjohn Institute Working Paper No. 13-198). Upjohn Institute for Employment Research. <https://doi.org/10.2139/ssrn.2281978>

Bassok, D., Dee, T. S., & Latham, S. (2019). The effects on incentives in early childhood education. *Journal of Policy Analysis and Management*, 38(4), 838-866.  
<https://doi.org/10.1002/pam.22149>

Benzies, K., Mychasiul, R., Kurilova, J., Tough, S., Edwards, N., & Donnelly, E. (2014). Two-generation preschool programme: Immediate and 7-year-old outcomes for low-income children and their parents. *Child & Family Social Work*, 19(2), 203-214. <https://doi.org/10.1111/j.1365-2206.2012.00894.a>

Bleach, J. (2015). Improving numeracy outcomes for children through community action research. *Education Action Research*, 23(1), 22-35.  
<https://doi.org/10.1080/09650792.2014.994016>

Boller, K., Paulsell, D., Del Grosso, P., Blair, R., Lundquist, E., Kassow, D. Z., & Raikes, A. (2015). Impacts of a child care quality rating and improvement system on child care quality. *Early Childhood Research Quarterly*, 30, 306-315.  
<https://doi.org/10.1016/j.ecresq.2014.10.001>

Bowne, J. B., Yoshikawa, H., & Snow, C. E. (2016). Experimental impacts of a teacher professional development program in early childhood on explicit vocabulary instruction across the curriculum. *Early Childhood Research Quarterly*, 34, 27-39. <https://doi.org/10.1016/j.ecresq.2015.08.002>

Bryant, D. M., Burchinal, M., & Zaslow, M. (2011). Empirical approaches to strengthening the measurement of quality: Issues in the development and use of

quality measures in research and applied settings. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality measures in early childhood settings* (pp. 33-47). Paul H. Brooks Publishing Company.

<https://doi.org/10.1080/10409289.2015.1036348>

Build Initiative. (2019). *Strong foundation for our youngest children*.

<https://www.buildinitiative.org>

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi experimental designs for research*. Houghton Mifflin.

Caronongan, P., Kirby, G., Malone, L., & Boiler, K. (2011). *Defining and measuring quality: An in-depth study of five child care quality rating improvement systems*. Office of Planning, Research, and Evaluation.

[https://www.acf.hhs.gov/programs/opre/cc/childcare\\_quality/five\\_childcare/five\\_childcare.pdf](https://www.acf.hhs.gov/programs/opre/cc/childcare_quality/five_childcare/five_childcare.pdf)

Cecconi, L., Stegelin, D. A., Pintus, A., & Allegri, R. (2014). Utilizing a global environmental assessment tool to facilitate professional development: The voices of kindergarten teachers in Italy. *International Journal of Early Childhood*, 46, 205-229. <https://doi.org/10.1007/s13158-014-0106-9>

Child Trends. (2013). [Home page]. <https://www.childtrends.org>

Cicconi, M. (2014, January). Vygotsky meets technology: A reinvention of collaboration in the early childhood mathematics classroom. *Early Education Journal*, 42(1), 57-65. <https://doi.org/10.1007/s10843-013-0583-9>

- Cloney, D., Cleveland, G., Hattie, J., & Tayler, C. (2016). Variations in the availability and quality of early childhood education and care by socioeconomic status of neighborhoods. *Early Education and Development, 27*(3), 384-401.  
<https://doi.org/10.180/10409289.2015.1076674>
- Compendium of Quality Rating Systems and Evaluations. (2010).  
[https://www.acf.hhs.gov/opre/resource/compendium\\_of\\_quality\\_rating\\_systems\\_and\\_evaluations](https://www.acf.hhs.gov/opre/resource/compendium_of_quality_rating_systems_and_evaluations)
- Connors, M. C., & Morris, P. A. (2015). Comparing state policy approaches to early care and education quality: A multidimensional assessment of quality rating and improvement systems and child care licensing requirements. *Early Childhood Research Quality, 30*, 266-279. <https://doi.org/10.1016/j.ecresq.2014.05.006>
- Cortes, J., & Hallam, R. A. (2016). QRIS: Empowering family child care providers as leaders. *Young Children, 71*(4), 14-19.  
<https://jstor.org/dtsblr/ycyoungcvhildren.71.4.14>
- Curby, T. W., & Brock, L. L. (2013). Teachers emotional support consistency predicts children's achievement gains and social skills. *Early Education and Development, 24*(3), 292-309. <https://doi.org/1.1080/10409289.2012.665760>
- Doyle, O., McGlanaghy, E., O'Farrelly, C., & Tremblay, R. E. (2016). Can targeted intervention mitigate early emotional and behavioral problems? Generating robust evidence within randomized controlled trials. *PLOS One, 11*(6), 1-22.  
<https://doi.org/10.1371/journal.pone.0156397>



- Durand, J., Hopf, A., & Nunnenmacher, S. (2016). Potentials and challenges of video-based self-reflection for the professionalization of early childhood education and care professionals. *Early Child Development and Care, 186*(1), 23-41.  
<https://doi.org/10.1080/03004430.2015.1031124>
- Elicker, J., Ruprecht, K. M., Langil, C., Lewsader, J., & Anderson, T. (2013). Indiana Paths to QUALITY: Collaborative evaluation of a new child care quality rating and improvement system. *Early Education and Development, 24*, 42-62.  
<https://doi.org/10.1080/10409289.2013.736127>
- Epstein, A., & Wilhite, G. L. (2015). Teacher efficacy in an early childhood professional development school. *International Electronic Journal of Elementary Education, 7*(2), 189-198. <https://iejee.com/index.php/IEJEE/article/view/74>
- Esplin, J. A., Neilson, B. G., Berghout Austin, A. N., & Fronk, A. (2019). Self-Report QRIS: Challenges with validation. *Early Education & Development 30*(6), 817-834. <https://doi.org/10.1080/10409289.2019.1591045>
- Faria, A., Greenberg, A. C., Hawkinson, L. E., & Metzger, I. (2016, June). Development and implementation of quality rating and improvement systems in Northcentral Region states (REL 2016-143). Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northcentral.  
<http://ies.ed.gov/ncee/edlabs/projects/projes/asp?ProjectID=4522>
- Faria, A., Hawkinson, L. E., Greenberg, A. C., Howard, E. C., & Brown, L. (2015, March). Examining changes to Michigan's early childhood quality and rating

improvement system (QRIS). (REL 2015-029). Washington, DC: US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northcentral. <http://ies.ed.gov/ncee/edlabs/projects/projes/asp?ProjectID=355>

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.

<https://doi.org/10.3758/BF03193146>

Fenech, M., Sweller, N., & Harrison, L. (2010). Identifying high-quality centre-based childcare using quantitative datasets: What the numbers do and don't tell us. *International Journal of Early Years Education*, *18*(4), 283-296.

<https://doi.org/10.1080/09669760.210.531615>

Ferguson, K. (2016). Lessons learned from using shadowing as a qualitative research technique in education. *Reflective Practice*. *17*(1), 15-26.

<https://doi.org/10.1080/14623943.2015.1123683>

Fiene, R., Lahti, M., Fiene, R., Elicker, J., & Zellman, G. (2015). Approaches to validating child care quality and improvement systems (QRIS): Results from two states with similar QRIS type designs. *Research Quarterly*, *30*, 280-290.

<https://doi.org/10.1016/j.ecresq.2014.01.005>

Fleer, M., Hedges, H., Fleer-Stout, F., & Bich Hanh, L. T. (2018). Researcher intersubjectivity: A methodology for jointly building an interactive electronic early childhood quality involvement/rating scale. *International Journal of*

*Research & Method in Education*, 41(1), 69-88.

<https://doi.org/10.1080/1743727x.2016.1219982>

- Forry, A. N. D., Iruka, I., Kainz, K., Torquati, J., Susman-Stillman, A., Bryant, D., & Smith, S. (2012). Identifying profiles of quality in home-based child care, Issue Brief OPRE 2012-20. Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Fox, L., McCulloch, P., Caronongan, P., & Herrmann, M. (2019). Are ratings from Tiered Quality Rating and Improvement Systems valid measures of program quality? A synthesis of validation studies from Race to the Top-Early Learning Challenge States. (NCEE 2019-4001). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. <http://ies.ed.gov/ncee>
- Fundus, J. L. (2015). Exploring the relationships between success on evaluations of preschool programs and children's pre-academic progress. *Delta Kappa Gamma Bulletin*.
- Gerdes, J., & Jefferson, T. (2015, November). How a professional learning community changed a family child care provider's beliefs and practices. *Young Children*, 6(10), 8-13. <https://www.naeyc.org/yc>
- Gilbert, J. L., & Harte, H. A. (2013). Coaching to quality: Increasing quality in early care and education programmes through community-University partnership. *Child Care in Practice*. 19(4), 397-411. <https://doi.org/10.1080/13575279.2013.800468>

- Great Start to Quality. (2016). [www.greatstarttoquality.org](http://www.greatstarttoquality.org)
- Green, S. B., & Salkind, N. J. (2014). Using SPSS for Windows and Macintosh: analyzing and understanding data. Pearson Education.
- Guernsey, L., & Oschorn, S. (2011). Watching teachers work: Using observational tools to promote effective teaching in the early years and early grades. *Early Education Initiative*. <https://earlyed.newamerica.net>
- Gunning, P., White, J., & Busque, C. (2016). Raising learners' awareness through L1-L2 teacher collaboration. *Language Awareness*, 25(1-2), 72-88. <https://doi.org/10.1080/09658416.2015.1122022>
- Gynther, K. (2016). Design framework for an adaptive MOOC enhanced by blended learning: Supplementary training and personalized learning for teacher professional development. *The Electronic Journal of e-Learning*, 14(1), 15-30. <https://academic-conferences.org/ejournal.htm>
- Hallam, R., Hooper, A., Bargreen, K., Buell, M., & Han, M. (2017). A two-state study of family child care engagement in quality rating and improvement systems: A mixed-methods analysis. *Education & Development*, 28(6), 669-683. <https://doi.org/10.1080/10409289.2017.1303306>
- Hallam, R. A., & Bargreen, K. N. (2013). Quality in family child care settings: The relationship between provider educational experiences and global quality scores in a statewide quality rating and improvement system. *Journal of Research Education*, 27, 383-406. <https://doi.org/10.1080/02568543.2013.822950>

- Hawkinson, L. E., Faria, A., & Bouacha, N. (2017). Quality improvement efforts among early childhood education programs participating in Iowa's quality rating system. *Regional Educational Laboratory Northcentral*. 85 pages.  
<http://ies.ed.gov/ncee/edlabs/projects/project/asp?projectID=4512>
- Hegland, S. M., Aronson, S. S., Isbell, P., Neelon, S. B., Rous, B. S., & Krajicek, M. J. (2011). Measuring health-related aspects of quality in early childhood settings. In Zaslow, M., Martinez-Beck, I., Tout, K., & Halle, T. (Eds.) *Quality Measures in Early Childhood Settings* (135-160). Paul H. Brooks Publishing Company.
- Hemmeter, M. L., Hardy, J. K., Schnitz, A. G., Adams, J. M., & Kinder, K. K. (2015, November). Effects of training and coaching with performance feedback on teachers' use of Pyramid Model practices. *Sage Journals*, 35(3), 144-156.  
<https://doi.org/10.1177/0271121415594924>
- High Scope Educational Research Foundation. (2003). *Preschool program quality assessment*. HighScope Press.
- High Scope Educational Research Foundation. (2018). [www.highscope.org](http://www.highscope.org)
- Holloway, S. D., Kagan, S. L., Fuller, B., Tsou, L., & Carroll, J. (2001). Assessing child-care quality with a telephone interview. *Early Childhood Research Quarterly*, 16(2), 165-189. [https://doi.org/10.1016/S-885-2006\(01\)00094-1](https://doi.org/10.1016/S-885-2006(01)00094-1)
- Hong, S. L., Howes, C., Marcella, J., Zucker, E., & Huang, Y. (2015). Quality rating and improvement systems: Validation of a local implementation in LA County and children's school readiness. *Early Childhood Research Quarterly*, 30, 227-240.  
<https://doi.org/10.1016/j.ecresq.2014.05.001>

- Hooper, A., & Hallam, R. (2017). Exploring the relationship between global quality and group engagement in toddler child care classrooms. *Journal of Research in Childhood Education, 31*(2), 215-226.  
<https://doi.org/10.1080/02568543.2016.1273287>
- Hu, B. Y., Fan, X., & Leong, S. S. L. (2015). Why is group teaching so important to Chinese children's development? *Australian Journal of Early Childhood, 40*(1), 4-12. <https://doi.org/10.1177/186393911504000102>
- Iloabuchi, E. N., Abraham, N. M., & Afangideh, S. T. (2016). Management of teaching staff for quality education delivery in secondary schools in Abia State, Nigeria. *American Journal of Educational Research, 4*(8), 617-623.  
<https://doi.org/10.12691/education-4-8-7>
- International Coaching Federation. (2016). *2016 ICF Global Coaching Study*.  
<https://www.coachfederation.org>
- International Coaching Federation. (2017, October). *Core Competencies*.  
<https://www.coachfederation.org>
- Ishimine, K., & Tayler, C. (2013). Assessing quality in early childhood education and care. *European Journal of Education, 49*(2), 272-290.  
<https://doi.org/10.1111/ejed.12043>
- Isner, T., Tout, K., Zaslow, M., Soli, M., Quinn, K., Rothenberg, L., & Burkhauser, M. (2011). Coaching in early care and education programs and quality rating improvement systems (QRIS): Identifying promising features. *Child Trends*.

- Jao, L., & McDougall, D. (2016). Moving beyond the barriers: Supporting meaningful teacher collaboration to improve secondary school mathematics. *Teacher Development, 20*(4), 557-573. <https://doi.org/10.1080/13664530.2016.1164747>
- Jeon, L., & Buettner, C. K. (2015). Quality rating and improvement systems and children's cognitive development. *Child Youth Care Forum, 44*, 191-207. <https://doi.org/10.1007/s10566-014-9277-7>
- Jeon, L., Buettner, C. K., & Hur, E. (2014). Family and neighborhood disadvantage, home environment, and children's school readiness. *Journal of Family Psychology, 28*(5). <https://doi.org/10.1037/fam0000022>
- Kabilan, M. K. (2016). Using Facebook as an e-portfolio in enhancing pre-service teachers' professional development. *Australian Journal of Educational Technology, 32*(1), 19-31. <https://doi.org/10.14741.ajet.2052>
- Karoly, L. A., Zellman, G. L., & Perlman, M. (2013). Understanding variation in classroom quality within early childhood centers: Evidence from Colorado's quality rating and improvement system. *Early Childhood Research Quarterly, 28*(4), 645-657. <https://doi.org/10.1016/ecresq.2013.05.001>
- Konza, D., & Main, S. (2015). The power of pedagogy: When all else fails. *The International Journal of Learning: Annual Review, 22*, 9-29. <https://doi.org/10.18848/1447-9494/CGP/v22/48746>
- Labone, E., & Long, J. (2016). Features of effective professional development learning: A case study of the implementation of a system-based professional learning

model. *Professional Development in Education*, 42(1), 54-77.

<https://doi.org/10.1080/19415257.2014.948689>

Lambert, R., Gallagher, P. A., & Abbott-Shim, M. (2015). An evaluation of the intensity of mentoring: Child outcomes. *Early Childhood Development and Care*, 185(8), 1314-1330. <https://doi.org/10.180/03004430.2014.992426>

Landry, S. H., Clancy-Menchette, J., Eisenberg, N., de Villiers, J., Barnes, M., & Starkey, P. (2014). Enhancing early child care quality and learning for toddlers at risk: The responsive early childhood program. *Developmental Psychology*, 50(2), 526-541. <https://doi.org/10.1037/a0033494>

Liechty, J. M., Schull, I. P., & Liao, M. (2009, Fall). Facilitating dissertation completion and success among doctoral students in social work. *Journal of Social Work Education*, 25, 399-414. <https://doi.org/10.5175/JSWE.2009.200800091>

Linder, S. M., Rembert, K., Simpson, A., & Ramey, M. D. (2016). A mixed-methods investigation of early childhood professional development for providers and recipients in the United States. *Professional Development*. 42(1), 123-149. <https://doi.org/10.1080/19415257.2014.978483>

MacFarlane, K., Lakhani, A., Cartmel, J., Casley, M., & Smith, K. (2015). 'Accepting change and enjoying the range': Applications of the Circle of Change methodology with professionals who support early childhood educators. *Professional Development in Education*, 41(2), 329-343. <https://doi.org/10.180/19415257.2014.486816>



- Maski, K. P., Loddenkemper, T., Ba, S. A., Allread, E. N., Urion, D. K., & Leviton, A. (2014). Development and implementation of a quality improvement curriculum for child neurology residents: Lessons learned. *Pediatric Neurology*, *50*(5), 4522-457. <https://doi.org/10.1016/j.pediatrneurol.2013.12.019>
- McLeod, R. H., Hardy, J. K., & Grifenhagen, J. F. (2019). Coaching quality in pre-kindergarten classrooms: Perspectives from a statewide study. *Early Childhood Education Journal*, *47*, 175-186. <https://doi.org/10.1007/s10643-018-0899-5>
- Melnick, H., Ali, T. T., Gardner, M., Meier, A., & Wecksler, M. (2017, June 1). *Understanding California's Early Care and Education System*. <https://learningpolicyinstitute.org/product/understanding-californias-early-care-education-system-report>
- Mendive, S., Weiland, C., Yoshikawa, H., & Snow, C. (2016). Opening the black box: Intervention fidelity in a randomized trial of a preschool teacher professional development program. *Journal of Education Psychology*, *108*(1), 130-145. <https://doi.org/10.1037/edu0000047>
- Mohler, G. M., Yun, K. A., Carter, A., & Kasak, D. (2009). The effect of curriculum, coaching, and professional development on prekindergarten children's literacy achievement. *Journal of Early Childhood Teacher Education*, *30*(1), 49-68. <https://doi.org/10.1080/10901020802668068>
- Morrissey, A., & Nolan, A. (2015, May). Just another meeting? Investigating mentoring for early childhood teachers in Victoria. *Australian Journal of Early Childhood*, *40*(2), 40-49.

- National Association for the Education of Young Children. (2003). *Early childhood curriculum, assessment, and program evaluation: Building an effective accountability system in programs for children birth through eight* [Policy Statement]. <https://www.naeyc.org>
- National Association for the Education of Young Children. (2009). *Developmentally Appropriate Practices in Early Childhood Programs Serving Children from Birth through Age 8*. <https://www.naeyc.org/resources/position-statements/dap/contents>
- National Center on Child Care Quality Improvement. (2015). *Quality rating improvement guidebook*. <https://.qrisguide.acf.hhs.gov/files>.
- National Center on Early Childhood Quality Assurance. (2019). *QRIS Resource Guide*. <https://ecquality.acf.hhs.gov/>
- Neumann, S. B., & Wright, T. S. (2010, September). Promoting language and literacy development for early childhood educators: A mixed-methods study of coursework and coaching. *The Elementary School Journal*, 111(1), 63-86. <https://doi.org/10.1086/653470>
- Norris, D. J., & Guss, S. (2016). Low quality of basic caregiving environments in child care: Actual reality or artifact of scoring? *Journal of Research in Childhood Education*, 30(4), 513-5228.
- Oke, E. B. (2016). Influence of early childhood instructional supervision on caregiver's effectiveness in Federal Capital Territory Centers, Abuja, Nigeria. *International Journal for Cross-Disciplinary Subjects in Education*, 7(1), 2682-2692. <https://doi.org/10.2533/ijcdse.2042.6364.2016.0365>

- Ottley, J. R., Coogle, C. G., & Rahn, N. L. (2015). The social validity of bug-in-ear coaching: Findings from two studies implemented in inclusive early childhood environments. *Journal of Early Childhood Teacher Education*, *36*, 342-361.  
<https://doi.org/10.1080/10901027.1100146>
- Park, J., & Ham, S. (2016). Whose perception of principal instructional leadership? Principal-teacher perceptual (dis)agreement and its influence on teacher collaboration. *Asia Pacific Journal of Education*, *36*(3), 450-469.  
<https://doi.org/10.1080/02188791.2014.961895>
- Perlman, M., Zellman, G. L., & Le, V. (2004). Examining the psychometric properties of the Early Childhood Environment Rating Scale-Revised (ECERS-R). *Early Childhood Research Quarterly*, *19*, 398-412.  
<https://doi.org/10.1016/j.ecresq.2004.7.006>
- Powell, D. R., & Diamond, K. E. (2013). Implementation fidelity of a coaching-based professional development program for improving Head Start teachers' literacy and language instruction. *Journal of Early Intervention*, *35*(2), 12-128.  
<https://doi.org/10.1177/1053815113516678>
- QRIS National Learning Network. (2013). <https://qrisnetwork.org>
- Race to the Top-Early Learning Challenge. (2014).  
<https://2.ed.gov/program/racetothetop-earlylearningchallenge/index/html>
- Reinking, A. K. (2015, March). Increasing accountability measures for early childhood teachers using evaluation models: Observation, feedback, and self-assessment.

*Current Issues in Education*, 18(1), 1-10.

<http://cie.asu.edu/ojc/index.php/cieatasu/article/view/1362>

Rentzou, K. (2017). Using rating scales to evaluate quality early childhood education and care: reliability issues. *European Early Childhood Education Research Journal*, 25(5), 667-681. <https://doi.org/10.1080/1350293x.2017.1356599>

Ronfeldt, A., & Campbell, A. L. (2016). Evaluating teacher preparation using graduates' observational ratings. *Educational Evaluation and Policy Analysis*.

<https://doi.org/10.3102/0162373716649690>

Sabol, T. J., & Pianta, R. C. (2015). Validating Virginia's quality rating and improvement system among state-funded pre-kindergarten programs. *Early Childhood Research Quarterly*, 30, 183-198. <https://doi.org/10.1016/j.recesq.2014.03.004>

Sandstrom, H., Moodie, S., & Halle, T. (2011). Beyond classroom-based measures for preschoolers: Addressing the gaps in measures for home-based care and care for infants and toddlers. In Zaslow, M., Martinez-Beck, I., Tout, K., & Halle, T. (Eds.) *Quality Measures in Early Childhood Settings* (317). Paul H. Brooks Publishing Company.

Saracho, O. N. (2015). Historical and contemporary evaluations for early childhood programmes. *Early Childhood Development and Care*, 185(5), 1255-1267.

<https://doi.org/10.1080/03004430.2014.989675>

Schulman, K., Matthews, H., Blank, H., & Ewen, D. (2012). A Count for Quality: Child Care Center Directors on Rating and Improvement Systems. *Center for Law and Social Policy, Inc.* <http://www.clasp.org>

- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The HighScope Perry Preschool study through age 40*. (Monographs of the HighScope Educational Research Foundation, 14).
- Setodji, C. M., & Le, V. (2013). Using generalized additive modeling to empirically identify thresholds within the ITERS in relation to toddler's cognitive development. *Developmental Psychology*, 49(4), 632-645.  
<https://doi.org/10.1037/a0028738>
- Sheridan, S. M., Edwards, A. P., Marvin, A. A., & Knoche, L. L. (2009). Professional development in early childhood programs: Process issues and research needs. *Early Education and Development*, 20(3), 377-401.  
<https://doi.org/10.1080/104092802582795>
- Shidler, L. (2009). The impact of time spent coaching for teacher efficacy on student achievement. *Early Childhood Education Journal*. 36(5), 453-460.  
<https://Edsovi.00045998.200936050.0011>
- Smith, S., Schneider, W., & Kreader, J. L. (2010, October). Features of professional development and on-site assistance in child care quality rating improvement systems. *National Center for Children in Poverty*.  
<https://eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED579792>
- Stockall, N. S. (2014, Summer). When an aide really becomes an aid: Providing professional development for special education paraprofessionals. *Teaching Exceptional Children*, 46(6), 197-205.  
<https://doi.org/10.1177/0040059914537202>

- Susman-Stillman, A., & Banghart, P. (2011). Quality in family, friends, and neighbor child care. *National Center for Children in Poverty*.  
<https://researchconnections.org>
- Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melruish, E., Elliot, K., & Totsika, V. (2006). Capturing quality in early childhood through environment rating scales. *Early childhood Research Quarterly*, 21, 76-92.  
<https://doi.org/10.1016/j.ecrsq.2006.01.003>
- Teras, H. (2016). Collaborative online professional development for teachers in higher education. *Professional Development in Education*. 42(2), 258-275.  
<https://doi.org/10.1080/10415257.2014.961094>
- Tonyan, H. A., Nuttal, J., Torres, J., & Bridgewater, J. (2017). Engaging with quality improvement initiatives: A descriptive study of learning in the complex and dynamic context of everyday life for family child care providers. *Early Education & Development*, 28(6), 684-704. <https://doi.org/10.1080/10409289.2017.1305152>
- Torquati, J. C., Raikes, H., & Huddleston-Casas, C. A. (2007). Teacher education, motivation, compensation workplace support, and links to quality of center-based child care and teachers' intention to stay in the early childhood profession. *Early Childhood Research Quarterly*, 22, 261-275.  
<https://doi.org/10.1016/j.ecresq.03.004>
- Tout, K. (2013). Look to the stars: Future directions for the evaluation of quality rating and improvement systems. *Early Education and Development*, 24, 71-78.  
<https://doi.org/10.1080/10409289.2013.741.912>

- Vermeer, H. J., Groeneveld, M. G., Larrea, I., Van Ijzendoorn, M. H., Barandiaran, S., & Linting, M. (2010). Child care quality and children's cortisol in Basque Country and the Netherlands. *Journal of Applied Developmental Psychology, 31*, 339-347. <https://doi.org/10.106/j.ecrsq.appdev.2010.05.001>
- Vitiello, V. E., Moas, O., Henderson, H. A., Greenfield, D. B., & Munis, P. M. (2012). Goodness of fit between children and classrooms: Effects of child temperament and preschool classroom quality on achievement trajectories. *Early Education and Development, 23*, 302-322. <https://doi.org/10.1080/10409289.2011.526415>
- Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.  
<https://doi.org/10.1080/14092802065379>
- Wilson, K. P., Dykstra, J. R., Watson, L. R., Boyd, B. A., & Crais, E. R. (2012). Coaching in early education classrooms serving children with autism: A pilot study. <https://eric.ed.gov/contentdelivery/servlet/ERICServlet?accno-ED535936>
- Yazejian, N., & Iruka, I. U. (2015). Associations among tiered quality rating and improvement system supports and quality improvements. *Early Childhood Research Quarterly, 30*, 255-265. <https://doi.org/10.1016/j.ecresq.2014.05.005>
- Zan, B., & Donegan-Ritter, M. (2014). Reflecting, coaching, and mentoring to enhance teacher-child interactions in Head Start classrooms. *Early Childhood Education Journal, 42*, 93-104. <https://doi.org/10.1007/s10643-013-0592-7>
- Zaslow, M., Tout, K., & Halle, T. (2011). Differing purposes for measuring quality in early childhood settings. In Zaslow, M., Martinez-Beck, I., Tout, K., & Halle, T.

(Eds.) *Quality Measures in Early Childhood Settings* (389-410). Paul H. Brooks Publishing Company. <https://brookespublishing.com/quality-measurement-in-EC>

Zellman, G. L., & Fiene, R. (2012). *Validation of quality rating and improvement systems for early care and education and school-age care* (Research-to-policy, research to practice brief OPRE 2012-29). Government Printing Office.

<https://www.acf.hhs.gov/>



### Appendix A: Quality Rating Improvement System Assessment Indicators

*Below is a list of some program quality indicators measured in the QRIS for the present study. The validation score represents the structural quality score. The Program Quality Assessment indicators represent the process quality score.*

#### Validation Indicators

| <i>Domain</i>  | <i>Subdomains</i>  |
|--|--|
| <i>Administration and program management</i>             | <i>Benefits, written policies, staff evaluation, sliding pay scale</i>                             |
| <i>Curriculum</i>  | <i>Daily routine, cultural competence plan, screening and assessments, consistent caregiving</i>   |
| <i>Environment</i>                                       | <i>Safety, staff ratios, outdoor time, nutrition plan, health records</i>                          |
| <i>Family and community partnerships</i>                 | <i>Communication with families, family engagement plan, community partnerships</i>                 |
| <i>Staff qualifications and professional development</i> | <i>Administration, lead teachers, assistant teachers, professional development hours and types</i> |

## Program Quality Assessment Indicators

| <i>Domain</i>                             | <i>Sub-Domains</i>   |
|---|--|
| <i>Learning Environment</i>               | <i>Safe and healthy, defined interest areas, outdoor space, varied and open-ended materials</i>  |
| <i>Daily Routine</i>                      | <i>Consistent daily routine, parts of the day, appropriate time for each part of the day</i>   |
| <i>Adult-Child Interactions</i>           | <i>Handling separation from home, support for child communication, encouragement of child initiatives, encouragement for peer interactions</i> |
| <i>Curriculum Planning and Assessment</i> | <i>Curriculum model, team teaching, comprehensive child records, anecdotal note taking process and procedures</i>                              |