Old Dominion University
ODU Digital Commons

Summer 2013

Quality Profiles in Early Childhood: An Example From Virginia's Quality Rating Improvement System

Kathryn M. Squibb
Old Dominion University

Recommended Citation

QUALITY PROFILES IN EARLY CHILDHOOD: AN EXAMPLE FROM

VIRGINIA'S QUALITY RATING IMPROVEMENT SYSTEM

by

Kathryn M. Squibb, M.S.Ed.
B.A., May 2003: College of William & Mary
M.S.Ed, August 2007: Old Dominion University

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirement for the Degree of

DOCTOR OF PHILOSOPHY

EARLY CHILDHOOD EDUCATION

OLD DOMINION UNIVERSITY
August 2013

Approved by:

_____
Andrea DeBruin-Parecki (Director)

_____
Angela Eckhoff (Member)

_____
Sabra Gear (Member)

ABSTRACT

QUALITY PROFILES IN EARLY CHILDHOOD: AN EXAMPLE FROM VIRGINIA'S QUALITY RATING IMPROVEMENT SYSTEM

Kathryn M. Squibb
Old Dominion University, 2013
Director: Dr. Andrea DeBruin-Parecki

Quality in early childhood settings has emerged as an important factor in determining whether the potential benefits of educational experiences before kindergarten will be realized. Research demonstrates that in order for such interventions to be beneficial to young children's development, the quality of their educational environments and experiences must be high. Quality rating and improvement systems (QRISs) are a method to assess, communicate, and improve the level of quality in early childhood settings. These rating systems have the potential to safeguard public investment by making programs accountable for quality, meet a consumer need for families seeking high quality care, and many incorporate support for programs to improve the levels of quality in a variety of early childhood settings. Across the country, states have developed QRIS programs in consultation with multiple stakeholders and they vary widely in terms of their structure and administration. However, little research has been conducted on the ability of these systems to deliver the intended benefits for children, families, and early childhood systems. The purpose of this study is to better understand the makeup of quality among the childcare programs participating in Virginia's pilot QRIS, and learn more about how the nature of quality in childcare settings associates with the overall program ratings assigned as part of participation in Virginia's QRIS. Results revealed a pattern of four distinct quality profiles among the participating programs using latent

profile analysis. Furthermore, two of these identified profiles were associated with the Star Ratings assigned by the comprehensive rating system, while the other two profiles showed no relationship to the overall Star Ratings.

This dissertation is dedicated to my parents, Don and Debbie, for their continued love and support of all I do.

vi

AKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Background

Problem statement. Preschool attendance prior to kindergarten entry has

received increased attention from researchers and policymakers, with over 40% of four

year olds in America attending public preschool, and another 32% attending private child

care programs (Barnett, Carolan, Fitzgerald, & Squires, 2012). Changes to the social

landscape in America have led to more single parent families and more mothers working

outside the home, both of which are contributing factors to the increased demand for

nonparental care of children prior to kindergarten entry, and increases in the availability

of public prekindergarten options (Halle, Martinez-Beck, Forry, & McSwiggan, 2011).

Research in the area of brain development has led to a better understanding of the

significant impact that children's experiences in their earliest years have upon their

success upon arrival at school, and for positive outcomes throughout childhood and into

adulthood (Shonkoff & Phillips, 2000). For these reasons, researchers and policymakers

have identified early childhood education (ECE) programs as having potential for helping

ameliorate effects of poverty and other risk factors to help children be better prepared for

school, ideally eliminating achievement gaps prior to school entry, and promoting better

outcomes across the lifespan (e.g., Burchinal & Cryer, 2003; Snow, Burns, & Griffin,

1998; Vandell, Belsky, Burchinal, Steinberg, & Vandergrift, 2010), and research from

seminal preschool intervention studies have documented the potential impact that

preschool experiences can have on children from disadvantaged backgrounds (Campbell

& Ramey, 1994; McCarton et al., 1997; Schweinhart, Barnes, Weikart, Barnett, & Epstein, 1993)

As a result, there has been increased investment in expanding preschool experiences for children prior to school entry, not only to account for the increased demand for nonparental care, but as a specific intervention to improve outcomes for young children as they enter school, resulting in reduced retention and intervention costs. These initiatives have experienced support from multiple stakeholders interested in improving economic outcomes for communities such as a stronger future workforce, lower crime rates and prison costs, as well as higher tax revenue and more productive citizens (Currie, 2000; Karoly, Kilburn, & Cannon, 2005). Thirty-nine states have implemented a state-funded prekindergarten program to supplement federal initiatives such as Head Start (Barnett et al., 2012). Still, large numbers of children arrive at school unprepared for cognitive, language, and social expectations of kindergarten, and significant achievement gaps are present before children begin formal schooling (Halle et al., 2011; Snow et al., 1998).

Quality in Early Childhood Education. Although early childhood education has been identified as having the potential to promote school readiness and future academic success (e.g., Burchinal & Cryer, 2003), most preschool programs do not provide the level of quality necessary to promote these types of outcomes, with the majority of programs for children aged three- to five-years old being described as mediocre, and programs for infants and toddlers characterized as mediocre to poor (Peisner-Feinberg & Burchinal, 1997).

As a result, researchers have turned their attention to learning more about the features of ECE programs which best promote outcomes for children and therefore protect investment in such programs. The term *quality* in ECE research is used to refer to various features or combinations of features that are thought to contribute positively to children's learning and development. Although this term is defined in various ways depending on the authors, stakeholders, or purposes of the definition, there is some commonality in describing different types of quality, and some consensus around how policymakers define *quality* as they set standards for early childhood education experiences which are publicly funded.

Although there is some controversy regarding how to measure quality, there is some consensus regarding the two subtypes of quality: *structure* and *process* (e.g., Cassidy et al., 2005; Phillips & Howes, 1987; Vandell & Wolfe, 2000). *Structure* refers to features of a program which are easily regulated, with some definitions adding that these are features which indirectly benefit children, such as teacher education level, wages, or teacher-child ratio (e.g., Peisner-Fienberg & Yazejian, 2010). Structural features of quality are usually included in licensing regulations as a minimum for operating a child care center or preschool, in part because they are easily monitored.

Process quality indicators are those features of a program which impact children directly, such as the language and interactions in a classroom or the learning materials that are used with children (e.g. Vandell & Wolfe, 2000). Process quality reflects the experiences that children have when they spend time in an early childhood setting, meaning that measurement is more dynamic and in-depth, typically requiring the use of

observational measurement tools (e.g., Peisner-Fienberg & Yazejian, 2010). See Chapter 2 for a more in-depth discussion of structural and process quality.

The desire to ensure that investments in ECE are likely to lead to positive outcomes for children has led researchers to investigate new lines of inquiry: What exactly is the nature of the relationship between structural and process quality? What levels of quality do children currently have access to? What are the thresholds of quality that are necessary in order to address the achievement gap? How can we most effectively improve levels of quality?

The desire to know more about quality levels in early care and education settings, combined with increased accountability for public funding has led to the implementation of quality rating improvement systems in several states.

Quality Rating Improvement Systems. Since the 1990s, state governments began to experiment with initiatives designed to assess the quality of ECE settings available to their communities (Mitchell, 2005). Quality rating improvement systems (QRISs) are a method to assess, communicate and improve the levels of quality available in ECE settings (Mitchell, 2005). Similar to rating systems in other industries (such as hotels or restaurants), QRISs assign rankings to childcare programs according to a specific set of standards, unique to the preferences and priorities of the state where the QRIS is developed. Chapter 2 provides an outline of which quality components are more commonly included in QRIS initiatives.

These systems emerged as states realized that their ECE program licensing standards represented a minimum threshold of health and safety, and that although accreditation through professional organizations such as NAEYC would ensure high

quality standards, it was not likely that the majority of programs would have the capacity or resources to reach that high standard (Westervelt, Sibley, & Shaack, 2008). Thus, QRISs were introduced as a means to assess quality along a broad continuum, and rate programs according to the level of quality provided to children in attendance.

As of 2009, 26 states have implemented QRIS initiative, and even more are piloting similar initiatives (Tout, Starr, Soli, Moodie, Kirby, & Boller, 2010). Although states frequently build on the progress or experiences of others, these systems are developed and operate independently of one another, are state-specific, and contain indicators of quality identified as a priority by the individual state where the system will be in practice. Chapter 2 provides tables of the makeup of QRISs throughout the United States, both in terms of the components of quality measured in their rating systems and the frameworks used to calculate comprehensive ratings for early childhood programs.

QRISs are usually developed as market-driven initiatives, based on a logic model indicating that rating and publicizing the levels of quality available for children would lead parents to make more informed choices about what ECE they select for their children and in doing so, higher quality offerings would flourish, while lower quality options would be driven out of the market (Zellman & Perlman, 2008).

Few states have been able to document the ability of QRISs to function in this market-based capacity (Schaack, Tarrant, Boller, & Tout, 2012). Unfortunately, parents' decisions in selecting childcare programs are rarely driven primarily by the quality of care offered by ECE programs. Rather, parents are more likely to use tuition cost and convenience of program location or hours to make decisions about where to enroll their children (Helburn, 1995), and may not have enough information or resources to be able to

demand higher quality options (Helburn & Bergmann, 2002). Further, states have not

sufficiently allocated funding for making parents aware of ratings and how to use them to

make decisions about childcare settings (Paulsell, Tout, & Maxwell, 2013). Another

barrier to QRISs' ability to function as a market force is that there are far fewer high

quality options available in areas where poverty is high, meaning that parents would not

be able to choose high quality options even if quality was their primary criterion in

selecting an ECE program (Helburn, 1995; Kontos, Howes, Shinn, & Galinsky, 1995;

Peisner-Feinburg et al., 2001).

Alternatively, QRIS initiatives are commonly intended to function as an important

link within a systems approach to early childhood supports and services available for

families intended to best prepare children for school readiness (Mitchell, 2009; Schaack

et al., 2012). From this systems-based approach, QRISs can offer important information

about the degree to which ECE programs are able to achieve the goals of preparing young

children for school entry, and potentially narrowing the earliest achievement gaps. Using

information from QRISs to evaluate the impact of ECE programs is a high-stakes

endeavor because of pressure from policymakers and other stakeholders under pressure to

demonstrate outcomes in a short timeframe (Swenson-Klatt & Tout, 2011; Zellman &

Perlman, 2008). These decisions are frequently taking place in an economic climate that

prioritizes the importance of funding only evidence-based and highly effective programs

with demonstrated positive outcomes for children and families.

Further, many states have linked QRIS ratings to a system of tiered childcare

subsidies, whereby states offering higher levels of quality (as defined by the particular

state's QRIS) receive higher rates of reimbursement from the government for enrolling

children who qualify for this subsidy. Since the provision of higher levels of quality costs more than substandard care (Peisner-Feinberg & Burchinal, 1997), the link to tiered childcare subsidies is critical for encouraging quality improvement within an ECE system. However, the high-stakes associated with such linkages highlight the vital importance of selecting components of quality and designing a QRIS structure which delineates meaningfully between levels of quality. For example, if various tiers within the system are reimbursed differentially for providing increasing levels of quality, it is imperative that the cut-off points between tiers are related to differences in quality that warrant different allocation of resources.

QRIS Evaluation. There has been a recent focus on establishment of quality rating improvement systems (QRISs) from the federal government. This is accompanied by a call for states with existing quality rating systems to demonstrate that they provide accurate information about the landscape of childcare quality available for the populations they serve (U.S. Department of Education & U.S. Department of Health and Human Services, 2011). QRISs are relatively new, and because they have grown out of policy initiatives rather than research pursuits, there is a sparse research base regarding the effectiveness of these systems for their stated goals and a lack of understanding about how different components fit together to create a comprehensive "snapshot" of quality presented in a rating (Tout, Zaslow, Halle, & Forry, 2009; Zellman & Fiene, 2012; Zellman, Perlman, Le, & Setodji, 2008).

Although states regularly select components for their QRISs that are research-based, and attempt to assign ratings in ways that they believe will reflect children's learning in meaningful ways, few states collect consistent, state-wide data about

children's learning and development in ways that can be used to inform these pursuits. Further, although there is a considerable research base regarding how individual components of quality contribute to children's learning, there is very little research to inform decision-making around the different methods that can be used to derive overall quality ratings from these individual components (Lugo-Gil et al., 2011).

Validating that the structure of a QRIS meaningfully differentiates levels of quality is an important step in establishing such systems as evidence-based and targeted to improve parent decision making, improving availability of high quality options, and ultimately using these systems for the intended purpose of improving school readiness for children. "Perhaps the most compelling way to define the importance of a validation study is that it can quantify if the quality ratings actually mean something important to programs, parents and children" (Lugo-Gil et al., 2011, p. 82).

Differentiating meaningfully implies that the tiers of quality identified by a QRIS reflect differences in quality that are also demonstrated through other means. Validating a QRIS system is a process of justifying the decisions about how tiers are delineated by demonstrating that they link to another method of differentiating quality. Zellman and Fiene (2012) emphasize validation's role in providing information about how well a QRIS' structure is functioning, and define validation for QRIS as, "a multi-step process that assesses the degree to which design decisions about program quality standards and measurement strategies are resulting in accurate and meaningful ratings" (p. 4).

Four approaches to validation activities exist. It is not mandatory to conduct activities in each area, nor will any specific validation activity allow a state to confirm

that their QRIS is valid. However, these four elements are presented in a recommended order of increasing intensity:

1. Examining the validity of the underlying components of a QRIS

2. Evaluating the measurement strategies and psychometric properties of the tools and data-gathering approaches used in the QRIS

3. Assessing the outputs (quality ratings) of the QRIS

4. Examining how the comprehensive ratings are related to children's outcomes (Zellman & Fiene, 2012)

Findings from validation activities may be used to help stakeholders make decisions when planning to implement a new QRIS initiative, or when states are considering revisions to a system already in place. However, it is important to note that this is a wide category of research activities, rather than a prescribed method, and that different validation activities will yield different aspects of information about a QRIS's effectiveness. "QRIS validation is not an effort that will result in a 'yes-no' decision about a QRIS" (Tout & Starr, 2013, p. 4). Instead, validation itself is a continuous improvement process.

The fourth approach to validation and the "gold standard" for validating a QRIS is establishing a link to child outcomes, or demonstrating that the tiers identified by the system are associated with differences in children's learning, development, and readiness for kindergarten. Practically speaking, child outcome evaluations are often prohibitively expensive, especially for state initiatives operating within tight budget constraints. Although demonstrating effects on children's learning is a necessary step in the process of fully establishing a QRIS, it is often prudent for states to do preliminary evaluation

work such as ensuring that the system is being implemented with fidelity and consistency. It is also recommended to engage in less costly validation work first, to ensure that the system itself reflects meaningful differences in the quality landscape among participating programs (Tout & Starr, 2013; Zellman & Fiene, 2012).

Virginia's QRIS. The Commonwealth of Virginia has taken great care to select components for their QRIS that are linked to literature on quality in childcare. The system is composed of four components of quality, including both structural and process quality. In an effort to develop an evidence-informed system, Virginia has identified teacher-child interactions, (a process quality component with strong ties to child outcomes) as a priority for the state by structuring its QRIS to reflect this priority. Teacher-child interactions have a strong research base of being linked to preacademic and social-emotional outcomes for preschool aged children, and therefore this aspect of quality was considered especially important (Howes et al., 2009; Mashburn et al., 2008).

Virginia constructed a points-based system as the framework for its QRIS, and each quality component has the potential to contribute a certain number of points towards any ECE program's total score. To reflect its emphasis on teacher-child interactions, for example, this component is worth the highest number of points towards the program's total score, and comparatively more than other components such as ratio, teacher education, or environmental quality (Kirby, Boller, & Tout, 2010).

Virginia's state initiative has made some strategic decisions related to how the system should be structured in order to increase the likelihood that it will, in fact, reflect levels of quality that are meaningful for children. However, these decisions have not been evaluated as they relate to the scores of programs who are actually participating in the

initiative. There are several ways to examine the relationships between the nature of quality among participating centers and the comprehensive quality ratings that are assigned through the QRIS. As mentioned above, one way to demonstrate that tiers of a QRIS are meaningful is to show that children attending differently rated program learn and develop at different rates. However, prior to taking on cost-intensive evaluations involving collection of child data, investigators may use other methods to learn more about the nature of quality in these programs and examine the relationship with ratings assigned by the initiative (Lugo-Gil et al., 2011).

One validation method would be to collect new data and conduct another measure of quality in participating programs, to determine whether there is an association with the comprehensive quality ratings assigned by the QRIS. Alternatively, it is possible to use statistical techniques to gain insight about the nature of quality across childcare programs using data that have been collected through the course of piloting the QRIS initiative. In the current study, data will be analyzed to reveal distinct quality profiles in the participating programs, using data that have been collected during the pilot phase of the Virginia QRIS initiative.

Should distinct profile exist in the quality among participating centers, their relationship to the comprehensive quality ratings will be examined to determine whether there is a meaningful link between Virginia's QRIS structure as outlined by its tiers and the existing subtypes of quality in Virginia's participating programs. Further, examining patterns and profiles among ECE programs in Virginia's QRIS may yield important information regarding how structural and process quality components fit together to generate profiles of quality based on relative strengths and weaknesses in these areas.

Purpose Statement

The first purpose of this study is to identify and describe patterns or profiles of quality existing among ECE programs participating in Virginia's QRIS. The second purpose of this study is to examine the relationship between identified profiles and the comprehensive quality ratings assigned through Virginia's QRIS.

The overarching purpose which connects these two questions is to examine whether the tiers of Virginia's QRIS relate to meaningful differences in the quality of ECE programs (as described by identified quality profiles) participating in the initiative.

Research Questions

1. What patterns or profiles of quality exist among ECE programs participating in Virginia's QRIS?

    a. How are structural and process quality described by the existing patterns or profiles present among participating ECE programs?

2. Is there a relationship between identified profiles and the comprehensive quality ratings assigned by Virginia's QRIS?

    a. What is the relationship between identified profiles and the comprehensive quality ratings assigned by Virginia's QRIS?

Hypotheses

It is hypothesized that distinct profiles of quality will be identified through examination of the quality rating data that have been collected through the pilot phase of Virginia's QRIS. It is hypothesized that, if identified, certain profiles of quality will be more strongly associated with a higher comprehensive quality rating, as compared with other profiles of quality.

Significance of the Study

The framework of how QRISs generate comprehensive quality ratings from individual quality components is significant; these systems are being used for high stakes purposes such as evaluating the effectiveness of ECE systems and programs, as well as a foundation for allocating resources such as tiered childcare subsidy reimbursements. Moreover, according to the QRIS logic model, overall quality ratings will be used by parents in their childcare decision making process, having an impact on the early childhood education market and ultimately influencing the success or failure of individual ECE programs. In order to assure continued investment and sustainability of QRISs, leaders must be able to demonstrate the ability of such systems to identify quality in a way that links to child outcomes. Prior to this endeavor, it is critical to determine whether the structure of the system itself is capable of differentiating between quality levels in a meaningful way.

This study will look at another way to conceptualize the nature of quality in ECE programs by examining what patterns or profiles exist among the components of quality measured by Virginia's QRIS. Identifying profiles contributes to the research base on structure and process quality by yielding more information about how these two features of quality exist together in actual ECE programs. The identification of profiles also contributes meaningfully to Virginia's policy context by lending some validation information about whether and how Virginia's comprehensive quality ratings may be associated with these profiles.

This study falls into the third approach to validation using Zellman and Fiene's (2012) framework: assessing the outputs of the rating process. The methods used in this

study may provide other states with a low-cost means to conceptualize quality in their state's ECE programs, without collecting new data.

If it is the case that identified profiles are associated with QRIS ratings, this will lend credibility to the system as an indication that the QRIS ratings are linked to patterns of quality which exist in the sample of ECE programs participating in Virginia's QRIS pilot. If profiles exist but are not associated with overall quality ratings, then this is instructive for policymakers so that adjustments can be made prior to conducting child outcomes evaluations, avoiding the expenditure of funds on an evaluation which is not likely to yield accurate results. In order to eventually demonstrate that different QRIS ratings are associated with differences in children's learning, it is imperative that the system is structured such that the different tiers of the system are not arbitrary and instead relate meaningfully to differences in quality in ECE programs.

Delimitations of the Study

This study includes the 358 initial ratings of ECE programs collected during the pilot phase of Virginia's QRIS from June 2009 until June 2012. Although this encompasses every initial rating assigned by the initiative during this time, the pilot was only available to ECE programs located in communities implementing Virginia's QRIS pilot, and participation was voluntary. This study is limited only to center-based child care programs, and does not include home-based child care programs.

This study intends to examine patterns and profiles of quality among the data on quality in these ECE programs collected as part of the QRIS pilot. Although this information has the potential to contribute to validation of the system, it addresses only one aspect of validation.

The following chapter includes a review of the existing literature exploring the relationship between quality in ECE programs and the ways that quality is used for accountability in early childhood systems. This review will include measurement of quality, an explanation of quality rating improvement systems, a summary of research on evaluating QRISs, and information on Virginia's QRIS.

Definition of Terms

Early childhood education (ECE) programs are defined in this document as classroom-based (rather than family-home based) child care and early education settings for children prior to kindergarten entry. This may include publicly funded programs such as Head Start or public prekindergarten, or private childcare in small businesses, not-for-profit programs, or religiously-affiliated child care programs.

Quality rating improvement system (QRIS) is defined in this document as a method to assess, communicate, and improve the level of quality provided in ECE programs (Mitchell, 2005). Although this generic definition does not describe the complexity and variety of different types of QRIS initiatives, it describes the common purpose among these systems.

CHAPTER II

LITERATURE REVIEW ·

The literature review is organized into five sections: an introduction to the relationship between quality in ECE programs and the context of accountability in America, a discussion on measurement of quality, an explanation of quality rating improvement systems, a summary of research on evaluating QRISs, and information on Virginia's QRIS.

Introduction

In 2012, more than 74% of American four-year-olds were enrolled in an early childhood education (ECE) program (public prekindergarten, Head Start, or private child care; Barnett et al., 2012). Preschool promises to be a powerful intervention to improve achievement for children in the first years of school, and is associated with positive outcomes throughout the lifespan, particularly for children from disadvantaged backgrounds (e.g., Burchinal & Cryer, 2003; Schweinhart et al., 1993; Snow et al., 1998). Evidence suggests that positive early childhood experiences with nonparental care can improve both social and academic outcomes for children, and has ignited an expansion of publicly available ECE programs for young children, particularly those at risk for school failure due to socioeconomic risk factors (Vandell et al., 2010). Nevertheless, children's readiness for kindergarten entry remains an ongoing concern for educators, policy-makers, and researchers. Costs for elementary grade retention and academic or social intervention remain high, and societal concerns for workforce development and community well-being persist.

A focal issue for researchers has been to construct a better understanding of the mechanisms by which ECE programs impact young children's future trajectories. To realize the potential that ECE programs hold for positively impacting children and the communities where they live, it will be important to better understand children's experiences prior to school entry and how to maximize benefits from these settings. Some studies have focused on capturing the level of quality that is available to American children attending public programs, as well as other types of care (Dowsett, Huston, Imes, & Gennetian, 2008). La Paro and colleagues (2009) examined the quality of 730 early childhood classrooms across 6 states and found that only 20% fell into the "high" quality range using two common measures of quality for ECE settings (CLASS; Pianta, LaParo, & Hamre, 2008, and ECERS; Harms, Clifford, & Cryer, 2005).

Other studies have focused on the link between the quality of care and the impact that this has on children's academic and social outcomes. Burchinal, Kainz, and Cai (2011) conducted a meta-analysis of 20 reports from five large scale studies of ECE experiences examined the associations between quality in prekindergarten with child outcomes. Findings indicated that although there was a stronger relationship between quality and academic outcomes, compared with social outcomes, associations between measures of quality and positive impacts on children were only modest in size (Burchinal et al., 2011).

A particular emphasis has been placed on assessing the level of quality accessible to children in poverty, since these programs are usually established for the purpose of addressing the achievement gap that exists between children experiencing economic and social risk factors and their more advantaged peers. For example, programs serving

children who receive subsidy reimbursement based on income qualifications are more likely to be of low quality, as measured by observations of teacher-child interactions and environmental quality (Antle, Frey, Barbee, Frey, Grisham-Brown, & Cox, 2008). Programs located in neighborhoods with less well-educated mothers and fewer social supports had lower ratings of environmental quality (Burchinal, Nelson, Carlson, & Brooks-Gunn, 2008). Efforts to establish public programs to ameliorate these disadvantages with the provision of higher quality ECE programs are plagued by additional concerns. For example, although center-based ECE programs established for at-risk children are more likely to have better educated teachers, they are also more likely to have higher teacher-child ratios and class sizes (Dowsett et al., 2008). In general, programs serving children in poverty tend to be of lower quality, raising concerns about whether they will be able to achieve the goals for which they were established.

LoCasale-Crouch and colleagues (2007) used cluster analysis to uncover patterns of quality in a nationwide sample of classrooms assessed using the CLASS tool, identifying several profiles of quality on a continuum from "high overall" to "low overall" in both emotional support and instructional quality. Findings indicate that instructional quality was a particular challenge in public programs across the country, even in programs established for the purpose of promoting school readiness (LoCasale-Crouch et al., 2007).

Private childcare options represent a significant portion of ECE settings in the United States, and the level of quality offered in these settings is often constrained by the market rate for childcare in the community where the program operates. This compounds the issue for communities where poverty is high, because programs must keep tuition

affordable, meaning that teacher salaries are low, and budgets for training, learning materials, and improvements to the environment are small. Ideally, parents would seek the highest quality option for their children and lower quality settings would have fewer clients and therefore either improve their levels of quality or be driven out of business (Schaack et al., 2012).

Parents choose child care settings based on a variety of factors, prioritizing affordability and convenience of location and operating hours (Johansen & Leibowitz, 1996; Seo, 2003; Zinzeleta & Little, 1997). Burchinal and colleagues (2008) found that parents also prioritize comprehensive service provision and a strong home-school connection when choosing care for their children. However, parents rarely select care using the same criteria as those identified by the research literature as being linked to child outcomes.

Measuring Quality

Quality in ECE programs can be conceptualized in several ways, according to different stakeholders. Parents may value cost and convenience, while policymakers might be most concerned with the maximum return on investment in initiatives designed to lower costs of retention and intervention in the elementary years. Research in early childhood learning and development defines quality as the elements of ECE programs likely to impact children's learning and development, with high quality experiences being most likely to yield these impacts as children interact meaningfully with teachers and peers, a stimulating environment, and age-appropriate learning materials (Howes et al., 2009; Mashburn et al., 2008).

Research has defined two types of measurable quality in ECE settings: structure and process, which together are thought to combine to yield global or overall quality (e.g., Cassidy et al., 2005; Phillips & Howes, 1987; Vandell & Wolfe, 2000). *Structure* is generally considered to be the features of a program which are easily regulated, but distal to the child's daily experience in the classroom. These are features which indirectly benefit children, such as teacher education level, wages, or teacher-child ratio (Howes, Phillips, & Whitebrook, 1992; Peisner-Fienberg & Yazejian, 2010). Structural features of quality are usually regulated by licensing standards as minimum requirements for operating a child care center or preschool. For research purposes, these aspects of quality are easily defined and consistently measured in various types of ECE programs.

Process quality indicators are those features of a program which are proximal to children's everyday experiences in ECE programs (Cassidy et al., 2005; Phillips & Howes, 1987; Vandell & Wolfe, 2000). They impact children directly, and include language and interactions in a classroom with teachers and peers, as well as learning materials that are used with children. Process quality reflects the experiences that children have when they spend time in an early childhood setting, meaning that measurement is more dynamic and in-depth, typically requiring the use of observational tools (Peisner-Fienberg & Yazejian, 2010). Because of the direct impact on a child's experience at any particular ECE program, process quality tends to be more strongly linked to children's learning and development (Howes et al., 2009; Mashburn et al., 2008).

Structural quality can be measured by simply collecting information on a variety of indicators, including: teacher education, teacher-child ratio, group size, teacher wages,

and parent fees. Process quality can be measured using instruments designed to capture information about the ways that children interact with adults, peers, and materials in the preschool classroom. Some tools focus strictly on teacher-child interactions, such as the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008) or the Caregiver Interaction Scale (CIS; Arnett, 1989). Other tools are specific to a particular domain of learning, such as language or literacy development (Early Language and Literacy Classroom Observation, ELLCO; Smith, Brady, & Anastasopoloulos, 2008).

Measurement of structure and process is not always a black and white endeavor; the number and type of materials present in a classroom would be considered an aspect of structural quality because it is easily prescribed and monitored, but the ways that these materials are used with children is an aspect of process quality because of the direct impact on children's experiences (Cassidy et al., 2005). And, some measurements are more comprehensive, and include process features of the classroom such as interactions, but also structural features such as the number and type of materials. Examples of this type of tool include the Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms et al., 2005) and the Infant Toddler Environment Rating Scale-Revised (ITERS-R; Harms, Cryer, & Clifford, 2003). The Environmental Rating Scales are composed of multiple constructs, including information about the child's learning environment and materials, degree of access to materials, interactions between teachers and children as well as peer-to-peer interactions, supports for health and safety, and family relationships. Because the Environmental Rating Scales are observational assessments that require a trained assessor to observe the classroom for several hours, it is classified here as a measure of process quality. Vandell and Wolfe's (2000) definition of process quality

supports this classification: process quality is defined by "actual experiences that occur in child care settings, including children's interactions with caregivers and peers and their participation in different activities" (p. 3), and that measures of process quality may "combine experiences across several areas that include health and safety provisions, interactions with caregivers, and age-appropriate materials" (p. 3).

Although distinct conceptually, several studies have documented moderate relationships between measures of structural and process quality (Howes et al., 1992; NICHD Early Child Care Research Network, 1996; Whitebook, Howes, & Phillips, 1990). Two recent studies have also pointed toward a more integrated view of quality. Rather than considering structure and process as separate indicators, researchers have begun to consider that some aspects of process quality may capture the means by which a child is able to benefit from the structural features of quality in the early childhood classroom. For example, a child's interactions with teachers and peers may play a significant role in how much the child is able to gain academically and socially from a particular learning environment.

Dominguez, Vitello, Maier, and Greenfield (2010) examined how high process quality can effect changes in children's learning behaviors over the course of the preschool year. Findings indicate that classrooms with relatively high scores in certain domains of the CLASS instrument (an observational measure of classroom interactions) were associated with higher rates of development in children's learning behaviors such as initiative, curiosity, engagement, persistence, goal setting, reasoning, and problem solving. By focusing on these approaches to learning, the authors suggest that ECE program quality is contingent on children's ability to maximize their opportunities for

learning, a more child-centered approach. Downer, Booren, Lima, Luckner, and Pianta (2010) report on the development of a child-centered observational assessment to measure how children interact with resources available to them in the classroom. The inCLASS is an observational assessment designed to evaluate an individual child's interactions with adults, peers, and learning materials or activities in the classroom setting, another variation on how to conceptualize quality in early childhood classrooms.

LoCasale-Crouch and colleagues (2007) characterize ECE program quality by analyzing data from observations of process quality using CLASS to examine patterns between different aspects of quality within one measurement tool. Cluster analysis revealed patterns of quality in a nationwide sample of prekindergarten programs assessed using the CLASS tool. Five distinct "quality profiles" were derived using this statistical technique to describe the nature of existing quality in public programs. Although a small number of classrooms fell into the "high overall" or "low overall" profiles, there were three additional profiles representing a combination of high, moderate, or low quality in the emotional or instructional domains. "By examining the interconnectedness of specific areas of interest, cluster analysis provides a potentially richer understanding of these relationships" (LoCasale-Crouch et al., 2007, p. 5). This analysis method revealed a new perspective on how different aspects of quality are associated with one another in the field.

The five quality profiles identified by LoCasale-Crouch and colleagues (2007) were associated with child outcomes in an effort to determine how these unique profiles might predict outcomes for children (Curby et al., 2009). Analyzing individual profiles yielded information indicating that high emotional support (with any level of

instructional support) will predict social-emotional competence, and that the dimension of concept development (one of 4 dimensions within the domain of instructional support) is the highest predictor of pre-academic skills, regardless of score levels in other instructional support dimensions and overall emotional support scores. It is important to note, however, that overall, only the highest quality childcare settings have been associated with even modest outcomes.

The elusive nature of robust impacts on child outcomes has motivated researchers to conduct studies which better define thresholds of quality beyond which positive impacts on children's cognitive, language, and social-emotional development are possible. For example, analyses of scores on the CLASS tool have identified *minimum* thresholds for quality necessary to promote impacts on children's development, but also demonstrated that increased quality continues to increase child outcomes, without an identified upper threshold (Burchinal, Vandergrift, Pianta, & Mashburn, 2010). By identifying the minimum thresholds in areas of emotional support (score of 5 of 7) and instructional support (3.25 of 7), these findings can inform goals for program improvement and demonstrate the need for increased funding to raise quality to meet these minimums.

Measurement of quality in childcare and prekindergarten programs becomes more meaningful when associations can be made with outcomes for children. No longer are preschool programs' primary function to provide daycare or babysitting services for working parents; early care and education programs have demonstrated the potential to impact children's school trajectories, and over 2.8 billion dollars have been invested in public prekindergarten programs dedicated to positively impacting school readiness for

children in poverty (Barnett et al., 2012). In order to ensure that this investment is worthwhile, and that all children are receiving equitable services and educational opportunities, researchers have focused on a variety of child outcomes and the features of quality that are most likely to foster the most growth.

Many studies examining the link between preschool quality and positive impacts on children have been conducted through large scale evaluation efforts in public state-funded prekindergarten programs usually linked in some way to public school systems (e.g., Burchinal et al., 2008; Burchinal et al., 2010; Curby et al., 2009; Mashburn, 2008; Mashburn et al., 2008). It is worth noting that significant numbers of young children attend ECE programs in the private sector, and that very little is known about how quality relates to child outcomes in these settings.

Most of these investigations of public prekindergarten quality follow the same basic framework: measures of quality are used as predictors of various child outcome variables, while controlling for demographic or sample characteristics. The most common measures of classroom quality were the CLASS and the ECERS, although one study measured structural quality using the National Institute for Early Education (NIEER) recommendations for program structure and teacher education (Mashburn et al., 2008). NIEER's recommendations for standards of quality in early childhood programs are used by state policymakers and program administrators to establish guidelines for high quality programs, and NIEER publishes an annual report assessing states according to these recommendations (Barnett et al., 2012).

Child outcomes measures for these studies typically fall into two categories: social-emotional and pre-academic. Social-emotional outcomes have almost exclusively been gathered through teacher-report using learning behaviors surveys and valid, reliable tools such as the Student Teacher Relationship Scale (STRS; Pianta, 2001), a teacher-report measure used to evaluate the quality of relationships between teacher-student dyads on the dimensions of Closeness, Conflict, and Dependency. Pre-academic skills have been assessed primarily through direct child assessment in pre-literacy, mathematical thinking, and in one instance, teacher report using the Academic Rating Scale (ARS; Rock & Pollack, 2002). Teachers using the ARS are required to evaluate children on a five-point scale according to their skills and abilities on basic academic tasks such as "Reads simple books independently."

Studies connecting quality with outcomes yield some important avenues for inquiry and analysis. Mashburn and colleagues (2008) investigated associations between different measures of quality (structural quality, CLASS, ECERS) and a basic set of child outcomes, and determined that CLASS was the only tool which significantly predicted outcomes for children. Mashburn (2008) examined data analysis methods that can be used to further break down the prediction value of quality for child outcomes. By considering quality as a continuous rather than categorical variable, effects were not detectable. However, when quality was broken into categories or levels of quality, associations with child outcomes were significant, pointing toward the potential for meaningful thresholds above which quality matters for child outcomes.

Beyond the need to better understand quality thresholds and their potential to impact children's learning, there is a practical need to help programs raise the level of

quality they provide and to help parents understand the importance of the quality of care they select for their children. State licensing regulations represent the minimum threshold of quality for ECE programs, and generally focus on features which are easily measured, such as structural quality and health and safety standards. Although state regulatory bodies place a high level of importance on structural quality, a study aiming to document the link between the rigor of state regulations and higher levels of quality in practice was not able to confirm that states with high standards had child care programs with higher levels of quality (Phillips, Mekos, Scarr, McCartney, & Abbot-Shimm, 2000).

Quality Rating Improvement Systems

Quality rating improvement systems (QRISs) have become increasingly common across the country in response to the growing demand for accountability for investments in ECE systems. Because stakeholders are specifically interested in the level of quality in ECE settings, these systems have gained in popularity because of the belief that they may influence ECE programs to raise quality, and provide information to parents so that they may select higher quality options for their children. State governments have instituted quality rating improvement systems (QRISs) to assess, communicate, and improve the level of quality offered in child care settings within their state (e.g., Mitchell, 2005; Wesley & Buysse, 2010). These initiatives establish standards of quality for early childhood settings and assess the degree to which programs meet these standards through a leveled system that reflects a continuum of increasing quality. Five elements characterize QRISs (Tout et al., 2010): assessment of ECE programs on established programs standards, assignment of an overall composite rating of program quality, communication of these ratings to parents and other stakeholders, the provision of

improvement supports such as technical assistance and financial incentives such as increased subsidy reimbursement or grant awards.

Ideally, QRISs can serve an important role in the overall ECE system by creating a market force by which parents and other stakeholders begin to demand increased availability or accessibility to high quality options. In theory, ECE providers themselves will raise the level of quality they offer, or the quality rating information can be used as an impetus to fund quality improvement initiatives or raise standards of quality for publicly monitored ECE programs. With more high quality options available and parents selecting high quality settings more frequently, children should have increased exposure to high quality ECE experiences and this, in turn, would promote more positive outcomes for children (Tout et al., 2010; Zellman & Perlman, 2008). However, the research literature on QRISs ability to deliver on this promise is small. There is not a sufficient evidence base to guide states in knowing how incentives promote participation, the effectiveness of different approaches to quality improvement, the effects of supports to help programs move between tiers, or to help states decide which quality components and overall framework will effectively describe quality in a way that is meaningful for children's development (Paulsell et al., 2013).

QRISs are also thought to be a means for accountability, to safeguard investment in an effort to ensure that publicly funded early childhood programs are of the quality necessary to achieve the goal of promoting child development, and to provide a needed benefit to parents. Other benefits include potential professionalization of the field through standardization, and their role as part of a comprehensive ECE system of support and development for young children and families (Schaak et al., 2012). Although there is no

standard model for QRIS development, five elements have been identified as defining a

QRIS:

1. Quality standards (components of quality to be measured)

2. Assignment of ratings based on quality standards

3. Technical assistance or other quality improvement supports

4. Financial incentives (to encourage participation or support improvement)

5. Publication or dissemination of ratings to parents and other stakeholders

(Mitchell, 2005; Tout et al., 2010)

State initiatives vary widely with regard to the selection of components of quality

to be assessed, how comprehensive quality ratings are calculated, and the means by

which this information is communicated to parents, politicians, and other stakeholders.

Further variation exists in the approaches that states take to quality improvement. Despite

the lack of evidence base, QRISs are being initiated by states seeking to document and

improve the quality of the child care system serving children prior to public schooling.

"These state actions are driving QRIS research and evaluation- not the other way

around," (Paulsell et al., 2013, p. 270).

Another factor motivating states to implement a QRIS is funding. QRIS initiatives

were a central topic of focus in the recent Race to the Top- Early Learning Challenge

Grant, and their potential to play an important role in closing the achievement gap points

to the importance of better understanding and refining the tiers within these systems (U.S.

Department of Education & U.S. Department of Health and Human Services, 2011). In

this federal grant competition, states were incentivized to demonstrate that their tiered QRIS was linked to early learning standards (standards outlined by the state for what young children should know and be able to do), provided information on child care quality to parents, was embedded in the ECE system within the state, and outlined a plan for validation of the system's tiered structure.

States are developing these systems largely independently of one another, and while there is some commonality in terms of which quality components are selected, states vary widely in terms of how these components are measured and how overall quality ratings are calculated using this information. Wesley and Buysse (2010) state, "There are no federal policies to guide the development of a QRIS and little encouragement for states to aim for consistent QRIS standards or measurements" (p. 8).

According to a compendium developed by Tout and colleagues (2010), most states include or address licensing as the foundation of their system, quality of the learning environment, qualifications of teaching staff, school-family partnerships, administration, and accreditation. Relatively fewer states include curriculum and early learning, child assessments, and ratio or group size, specific issues of health and safety, diversity, provisions for children with special needs, and the programs involvement in the community. Thirteen different quality components were identified in the comprehensive review of 26 QRIS initiatives. Table 1 below outlines some of the most common quality components utilized in QRIS initiatives.

Table 1

Common QRIS quality components

	Licensing	Ratio and Group Size	Health and Safety	Curriculum	Environment	Staff Qualifications
Calif., LA	Yes	Yes			Yes	Yes
Colorado	Yes	Yes			Yes	Yes
Delaware	Yes			Yes	Yes	Yes
District of Columbia	Yes			Yes	Yes	Yes
Florida Miami-Dade	Yes	Yes		Yes	Yes	Yes
Florida, Palm Beach	Yes	Yes		Yes	Yes	Yes
Illinois	Yes				Yes	Yes
Indiana	Yes			Yes	Yes	Yes
Iowa	Yes	Yes	Yes		Yes	Yes
Kentucky	Yes	Yes			Yes	Yes
Louisiana	Yes	Yes			Yes	Yes
Maine	Yes			Yes	Yes	Yes
Maryland	Yes				Yes	Yes
Minnesota	Yes		Yes	Yes	Yes	Yes
Mississippi	Yes			Yes	Yes	Yes
Missouri	Yes			Yes	Yes	Yes
New Hampshire	Yes			Yes	Yes	Yes
New Mexico	Yes	Yes		Yes	Yes	Yes
North Carolina	Yes	Yes		Yes	Yes	Yes
Ohio	Yes	Yes		Yes		Yes
Oklahoma	Yes		Yes		Yes	Yes
Oregon	Yes	Yes				Yes
Pennsylvania	Yes		Yes	Yes	Yes	Yes
Tennessee	Yes	Yes			Yes	Yes
Vermont	Yes				Yes	Yes
Virginia	Yes	Yes			Yes	Yes
Total	26	13	4	14	24	26

Source: Tout et al. (2010)

States' selection of quality components for inclusion in QRIS initiatives are a reflection of the elements of quality that the state hopes to promote in its childcare programs (Paulsell et al., 2013). Frequently, this process is also driven by values and beliefs of various stakeholders involved with implementing the QRIS, as well as the political climate in which the initiative is launched.

The tiers of a QRIS initiative can be structured in several ways by deciding how individual components will be arranged to generate an overall quality rating score. There are three primary formats for composing these ratings: a building blocks system, a points-based system, or a combination of the two (a variation of a points-based system with built-in thresholds similar to the building block style). Building block systems specify that particular components of quality or levels within a component must be satisfied prior to being eligible for higher level ratings. In a points-based system, ECE programs "earn" points across several components of quality and these points are summed to create an overall score. Cut-off points are used to determine overall ratings (for example, a certain total number of points indicates a particular overall rating). States using a combination approach each have a different balance of building blocks and points elements. Table 2 describes each state's QRIS structure.

Table 2

QRIS structures

	Building blocks	Points	Combination/Other
California, LA Co.	X		
Colorado		X	
Delaware	X		
District of Columbia	X		
Florida Miami-Dade			X
Florida, Palm Beach			X
Illinois	X		
Indiana	X		
Iowa			X
Kentucky	X		
Louisiana			X
Maine	X		
Maryland	X		
Minnesota		X	
Mississippi	X		
Missouri			X
New Hampshire			X
New Mexico	X		
North Carolina		X	
Ohio	X		
Oklahoma	X		
Oregon			X
Pennsylvania	X		
Tennessee			X
Vermont		X	
Virginia		X	
Total	13	5	8

Source: Tout et al. (2010)

Points-based systems are limited by the fact that there are no minimums for quality within any particular component, meaning that programs with lower levels of quality in one area may compensate with higher quality in another area. This contributes

to higher variation within tiers compared with building block systems, where states can be assured that programs have achieved at least minimum levels of quality in particular component areas (Tout et al., 2010). In their compendium of QRISs, Tout and colleagues (2010) report that half of states with QRIS initiatives utilize a building blocks system, with only 5 (including Virginia) using a points-based system. The remainder have selected a combination of the two approaches, or have alternative structures in which they do not assign ratings, but use their tiers as a method to assess how far programs have raised quality above licensing standards.

There is much variability in terms of the relationships that these programs have with the childcare regulation bodies in these states, which are the organizations issuing and monitoring health and safety standards through licensing. Some states have tied the rate of child care subsidy reimbursements to the level of quality the program offers, in recognition of the fact that providing high quality care comes at a cost, and to help increase access to high quality settings for children whose families are eligible for child care subsidies. Of the 26 QRISs reviewed by Tout and colleagues (2010), most systems were linked to tiered reimbursement within the childcare subsidy program. States implementing QRIS programs also provide varying levels of support and incentive structures encouraging or rewarding high achievement, and vary in their provision of resources and technical assistance to improve the quality provided by participating centers (Peisner-Feinberg & Yazejian, 2010). Eleven states provide bonuses for programs for achieving certain levels of quality, either as a one-time bonus or on an annual basis for maintaining quality levels (Tout et al., 2010).

Evaluation of QRIS

Evaluation of efforts to rate and improve ECE program quality is a priority outlined by the Administration for Children and Families, Office of Planning, Research, and Evaluation (Zellman, Brandon, Boller, & Kreader, 2011; Zellman & Fiene, 2012). Of the states implementing QRIS initiatives, most are engaging in some type of evaluation activity, either internally or through an external evaluation contract. However, although evaluation activities are being prioritized, research designs used to date have not permitted causal conclusions regarding the impact that participation has on quality improvement (Paulsell et al., 2013).

Seven states are conducting validation studies of their tiered structures. Validation differs from evaluation in that evaluation seeks to demonstrate the impact or implementation of an initiative, whereas "the central question in a validation study is whether the different levels that make up a QRS represent different levels of quality," (Tout et al., 2010, p. 192). Validation activities are considered a preliminary step to evaluation of QRIS impact, because it would be impossible to describe the impact of different quality levels without first assuring that the tiers are accurately differentiating between quality levels (Zellman & Fiene, 2012).

Fewer states are conducting evaluations to link QRIS ratings to child outcomes. Although evaluation activities are becoming more common, states have struggled to demonstrate robust outcomes demonstrating their positive impact on children's learning and development (e.g., Tout et al., 2010). Instead, several states have focused on preliminary evaluation activities regarding implementation fidelity and validation as wise

investments prior to engaging in studies of child outcomes, to ensure that systems are well-established and logically sound before beginning high-stakes, expensive investigations involving child outcomes (Zellman & Fiene, 2012).

It is difficult to reconcile a robust literature base demonstrating the potential for influencing young children's trajectories through high quality ECE programs with small to nonexistent associations between quality levels and outcomes for children at school entry (Schaack et al., 2012). However, evaluating initiatives in a policy context is very different than evaluating quality improvement interventions specifically designed to impact children's learning and development. Research in the area of implementation science suggests that the gap between an evidence-based program and the scaled-up versions of these programs is large and can result in lackluster findings (Downer, 2013).

Although the logic model for QRISs point to the eventual outcome of improved outcomes for children, there are several mediating factors at work. These include market forces, parental decision making processes, and the context of quality improvements. Improvement depends on a series of systems-level changes, possibly including better higher education systems or training to improve teacher qualifications, more powerful links to subsidy to incentivize and promote high quality care, and changes in how people view themselves as members of the profession (Schaak et al., 2012). Because of the challenge of demonstrating QRIS's ability to impact child outcomes, many states are turning to evaluation of these additional outcomes to show QRIS's impact on the early childhood system overall.

There is a growing body of information and guidance regarding how to strategically engage in preliminary activities prior to evaluation, to increase the likelihood

that investigations will be able to accurately measure the impact of systems-level

initiatives such as QRIS. Zellman and Fiene (2012) outline four types of evaluation

activities, and suggest that the order in which these activities are conducted is extremely

important for the purpose of safeguarding investment and ensuring accurate results. First,

in selecting components for a QRIS, a state should take care to choose features of quality

which are tied to the research literature. In designing a system likely to reflect differences

in children's learning and development, it is critical to select variables, or components,

that are valid for this purpose. Second, states should plan for ongoing process evaluation

of implementation processes to ensure that the QRIS is being executed with fidelity to the

intended process and also that the system is being conducted consistently across

participants. If states are implementing QRIS with varying degrees of fidelity, or if the

degree of fidelity is unknown, this will negatively impact the system's ability to reliably

communicate comparable data between programs or communities within the state. The

third category of evaluation activities is validation of the structure of QRIS itself. This

refers to investigations seeking to ensure that the tiers of a QRIS meaningfully reflect

differences in quality for ECE programs. It is essential to demonstrate that the QRIS is

functioning in this most basic role, particularly if the tiers are intended to represent

differences in an ECE program's ability to impact children's learning and development.

The step of validating the structure of a QRIS is a critical precursor to engaging in the

fourth evaluation activity: linking to child outcomes. The ultimate outcome in Zellman &

Perlman's (2008) QRIS logic model is better outcomes for children.

In order to demonstrate this eventual outcome, it will be necessary for QRIS

initiatives to demonstrate that their ratings are related to differences in children's

learning. From a policy perspective, this represents a necessary step toward ensuring the sustainability of such programs because it is the foundation for families to use ratings with confidence to make decisions for their children. Validating the tiers also demonstrates the potential for QRIS ratings to accurately differentiate between quality levels, meaning that future findings from child outcomes studies may be attributed to the different types of early care and education children are receiving.

"Because the components of quality addressed through QRIS standards are likely selected in a political climate through consensus among diverse stakeholders about what quality means, it is critical to ask hard questions about whether a QRIS measures what it purports to" (Wesley & Buysse, 2010, p. 9). Decisions about quality components and rating structures are frequently made by stakeholders hoping to prioritize particular elements of quality, and the restrictive budgets of many states can result in rushed implementation or inappropriate expectations regarding the timeline under which outcomes will be achieved (Swenson-Klatt & Tout, 2011; Zellman & Perlman, 2008). Although QRIS initiatives have enjoyed popularity among states over the past decade, there is very little empirical research confirming their effectiveness to improve program quality or impact children's learning and development (Tout et al., 2009).

Studies attempting to demonstrate that QRIS participation results in increased quality for ECE programs have been primarily non-experimental, and have found largely non-significant results regarding the impact on ECE programs' quality (Barnard, Smith, Fiene, & Swanson, 2006; Cheatam, Pope, & Myers, 2005; Norris, Dunn, & Eckert, 2003; Zellman et al., 2008). There have not been any experimental studies to examine how QRIS participation impacts children's learning and development.

Only one study has attempted to correlate components of a QRIS to outcomes for children. A 2008 evaluation of Colorado's QRIS (Zellman et al., 2008) was designed to examine a link between the components of the system with academic and social outcomes in children attending participating ECE programs. However, this study failed to find a relationship between children's learning and either individual quality components or overall quality ratings. A major limitation in this study was the attrition rate of children in participating ECE programs; fewer than 10% of children originally selected as participants were still enrolled in their ECE programs by the end of the study. This highlights a considerable hardship for conducting child outcome research in the context of child care programs, and emphasizes the importance of preliminary evaluation activities.

Research on structure of QRIS initiatives is also rare, and very little is known about where to differentiate between tiers in meaningful ways that will be reflect in child outcomes, leaving states to determine for themselves how to delineate between tiers of quality, and how to define the level of quality within individual components. For example, although 20 different state QRISs include observations of environmental quality using ECERS, states vary widely in terms of what scores on this tool qualify a program for inclusion in the top tiers. The District of Columbia requires a score of 4 (on a scale of 1-7), Mississippi requires a 5.1, Florida requires a 5.5, and Kentucky requires a score of 6 (Tout et al., 2010). Decisions about how to measure other quality components such as curriculum implementation or teacher education and qualifications vary even more widely because there is not a valid, reliable tool on which to rely.

Similar decision-making processes take place regarding determining cut-off points between individual tiers of a QRIS structure. In lieu of a substantial evidence base suggesting thresholds of quality (across different components of quality) that are likely to differentially impact children's learning and development, developers of QRIS initiatives must make decisions on these issues based on their preferences and estimates about what is likely to promote positive outcomes for children.

However, if the structure of QRIS does not accurately reflect differentiations in quality that matter for young children's learning and development, either because the composition of the overall ratings masks the unique contribution of each component or because the cut-off points between tiers are arbitrary, it is unlikely that child outcome studies will be able to demonstrate positive impacts of QRIS participation.

Virginia's QRIS

In response to concerns in Virginia regarding school readiness at kindergarten entry and accessibility of high quality early childhood education, Virginia began developing a voluntary quality rating improvement system in 2007, overseen at the state level in a partnership between the Office of Early Childhood Development (OECD) and the Virginia Early Childhood Foundation (VECF) (Kirby et al., 2010). The author has been worked closely with VECF and VDSS on Virginia's QRIS since 2007, employed as a local administrator and as an independent consultant to facilitate implementation of the pilot and to examine and interpret data from the pilot years.

In 2009, pilot ratings began, and between June 2009 and June 2012, 358 initial quality ratings have been assigned. Some programs have been rated more than once, for a total of 419 total assessments conducted. Prior to 2009, the initiative operated a

preliminary phase during which programs received information and feedback on the quality of their ECE program, but overall ratings were not calculated, posted, and made available to parents and other stakeholders.

Although Virginia's QRIS is administered at the state level through the collaboration between OECD and VECF, much of the coordination of the initiative occurs locally (Kirby et al., 2010). Individual communities volunteer and are selected for participation in the QRIS, meaning that ECE programs not located in these communities cannot participate in QRIS. Between the time period of June 2009 and June 2012, 18 communities participated in Virginia's QRIS.

Programs are rated every two years, and all types of center-based ECE programs (e.g. nonprofit, private, public, and religiously affiliated) are eligible to participate, provided that they are in good standing with their regulatory authority (e.g., licensing, Virginia Preschool Initiative regulations, Head Start monitoring). Part of the funding for the QRIS is garnered by local entities, so levels of participation and the extent of support services and technical assistance vary regionally across the state. Communities are also largely responsible for the improvement phase of QRIS, hiring mentors and providing funding for quality improvements based on local fundraising and investment efforts (Kirby et al., 2010).

Virginia's QRIS uses a points-based system and measures quality across four components, each of which is worth a different number of points toward a cumulative score. Different components are weighted to emphasize the amount of contribution made toward the final overall score. Each program also receives an overall "Star Rating" according to the number of cumulative points earned across all four quality components.

This voluntary system was designed to be market-driven, and builds upon the health and safety standards for child care licensing in Virginia to indicate how far above licensing an ECE program has reached.

There are four components of quality measured by Virginia's QRIS: 1) education qualifications and training, 2) teacher-child interactions, 3) ratios and group size, and 4) learning environment (Virginia Star Quality Initiative, 2009). Each of these components is linked in the research to positive outcomes for children in either social or academic domains (Burchinal et al., 2011; Hamre & Pianta, 2005; NICHD, 2002); however, the extent to which they contribute to children's development varies both by component and across different investigations.

Information about teacher education and qualifications is submitted in an electronic form designed for this purpose at the beginning of the rating process. This self-report form is reviewed by the administrative hub and points are assigned based on the level of education and qualifications of all program staff, including directors, teachers, and assistant teachers. The maximum number of points that can be awarded in this standard area is 40.

Ratio and group size information is also submitted during the application process via self-report form, capturing information about teacher-child ratios over the period of one week. This document is reviewed and points are assigned according to the reported ratios and group sizes. The maximum number of points available in this area is 30.

Following the submission of required information on structural quality, an observer trained through the state is assigned to visit a center to gather information about process quality in a representative sample of classrooms. One of every three classrooms

in the toddler, three-, and four-year-old age ranges is observed. Observers select and use the appropriate version of the CLASS measure to assess teacher-child interactions: PreK CLASS (Pianta et al., 2008) for three- and four-year-old classrooms and Toddler CLASS (LaParo, Hamre, Pianta, 2011) for classrooms of children under the age of three. In the same visit, observers collect information using the appropriate Environment Rating Scale: ECERS (Harms et al., 2005) for children 30 months to five years old, or ITERS (Harms et al., 2003) for children under 30 months old (Kirby et al., 2010).

Observers report scores to the administrative hub for Virginia's QRIS, where administrators apply a rubric to the raw score to determine how many of the available points a program has earned. Teacher-child interactions are worth a maximum of 60 points, and learning environments are worth a maximum of 40 points (VSQI, 2009). Through the rating calculation process, classroom level observation scores are converted to program level scores by averaging across classrooms.

Overall quality ratings are determined based on the total number of points that a program earns across all four components, according to the following scale: 34-50 earned points is a 1-Star rating, 51-84 earned points is a 2-Star rating, 85-118 earned points is a 3-Star rating, 119-152 earned points is a 4-Star rating, and 153-170 earned points is a 5-Star rating. Across Virginia, all ratings have been between one and four stars, with no programs being assigned a 5-Star rating (Kirby et al., 2010). Quality components and weighting decisions were made in consultation with experts and in consideration of a review of the literature and best practices from other states (Virginia Department of Social Services, 2009). However, because decisions regarding delineation between tiers

of the system were made during the pilot phase, they were not driven by associations with child outcomes or relations to external measures of quality.

Recently, outcomes for children attending participating QRIS centers in Virginia have been evaluated. Sabol, Pianta, Downer and Cao (2011) compared star levels in 71 participating public prekindergarten programs with child outcome data on the Phonological Awareness Literacy Screening (PALS-PreK; Invernizzi, Sullivan, Meier, & Swank, 2004). Results indicate that attending a 3-Star or 4-Star rated program was associated with increased growth compared with attendance in a 2-Star program. This investigation utilized a complex set of controls for demographic information about participating children; however, differences among the samples of children attending differently rated programs remain a concern. Counterintuitively, children attending 2-Star rated programs began the school year with higher literacy skills than did their peers attending 3- and 4-Star rated programs. Additionally, this study considered only public prekindergarten settings, rather than the entire population of ECE programs participating in Virginia's QRIS, and it is unlikely that these programs demonstrated great variation on structural measures of quality due to the standards in place for public prekindergarten programs regarding quality features such as teacher qualifications and teacher-child ratio are monitored by funders. More information about the landscape of quality across all ECE programs participating in Virginia's QRIS is needed.

The third chapter will discuss the methodology of the current project, including a description of the purpose and research questions, research design, information on the population and sample under study, description of instrumentation, procedures used to collect data, and data analysis plan.

CHAPTER III

METHODOLOGY

This chapter is organized into the following sections: description of the purpose and research questions, research design, information on the population and sample under study, description of instrumentation, procedures used to collect data, and data analysis plan.

Purpose Statement

The first purpose of this study is to identify and describe patterns or profiles of quality existing among ECE programs participating in Virginia's QRIS. The second purpose of this study is to examine the relationship between identified profiles and the comprehensive quality ratings assigned through Virginia's QRIS.

The overarching purpose which connects these two questions is to examine whether the tiers of Virginia's QRIS relate to meaningful differences in the quality of, ECE programs (as described by identified quality profiles) participating in the initiative.

Research questions. The questions to be addressed in this study are:

1. What patterns or profiles of quality exist among ECE programs participating in Virginia's QRIS?

    a. How are structural and process quality described by the existing patterns or profiles present among participating ECE programs?

2. Is there a relationship between identified profiles and the comprehensive quality ratings assigned by Virginia's QRIS?

    a. What is the relationship between identified profiles and the comprehensive quality ratings assigned by Virginia's QRIS?

Research Design

To address the first research question, profiles, or subtypes of quality were derived by examining patterns in quality features in these programs. To address the second research question, the relationship between profiles of quality and the comprehensive ratings was examined.

In this quantitative, non-experimental study, latent profile analysis was used to determine whether there are patterns or profiles in the quality of participating ECE programs that can be identified using data that have been collected during the pilot phase of the initiative. Following the categorization of programs into subtypes, the relationship between profile and comprehensive quality ratings was examined to determine whether there is a significant correlation between Virginia's QRIS ratings and the identified profiles of quality in Virginia's participating programs.

This study falls into the category of research activities designed to contribute to the validation of QRIS initiatives. Specifically, this study falls under the third category of validation: assessing the outputs of the rating process (Zellman & Fiene, 2012), because it relates the comprehensive ratings to another means of differentiating quality between ECE programs.

Further, examining patterns and profiles among ECE programs in Virginia's QRIS yields descriptive information regarding how structural and process quality components fit together to generate profiles of quality based on relative strengths and weaknesses in these areas. Comprehensive information about ECE quality subtypes can inform policy decisions regarding technical assistance or allocation of funding to support quality improvement.

Population and Sample

The population under study is early childhood education (ECE) programs in the Commonwealth of Virginia, represented by the sample of programs who have volunteered to participate in the pilot of Virginia's QRIS.

Participants include the 358 ECE programs who have received comprehensive quality ratings as part of Virginia's QRIS from June 2009 until June 2012. This time period was selected because it marks the beginning of the state's entry of information into an electronic database; prior to this date, information on pilot data were available only in paper format.

Participants have received a total of 358 initial quality ratings during the selected time period. In addition to the 358 initial ratings assigned, Virginia's QRIS also assigned 73 re-ratings to programs whose initial ratings had expired. These re-ratings are excluded from the sample under study for the purpose of consistency.

These programs operate in 18 diverse communities across Virginia, in urban, suburban, and rural settings. Further, participating programs represent a variety of types of care, including small businesses, nonprofit organizations, corporate childcare, religiously affiliated or exempt programs, public pre-kindergartens, Head Start programs, accredited programs, and those accepting childcare subsidies.

Ratings assigned during this time frame range from the *1-Star* level to the *4-Star* level. No *5-Star* ratings had been assigned as of June 2012. Of the ratings assigned during this time period, there is one 1-Star rating, 77 2-Star ratings, 187 3-Star ratings, and 93 4-Star ratings. All 358 initial ratings will be considered in the analysis for this study.

Participating programs were recruited locally by the 18 communities across Virginia, and participation was voluntary. Incentives were provided for pilot programs at varying levels in each community by local administrators. These incentives ranged from minimal support and training to extensive financial and marketing benefits. Following the rating, programs received quality improvement mentorship according to guidelines set forth by the state hub.

Measures

Data were collected on the four quality components identified by Virginia's QRIS using a combination of self-report forms completed by the ECE program and observations conducted by individuals trained by the state for this purpose. Data on *structural* quality such as teacher education and qualifications; teacher-child ratio and group size were collected via self-report forms. Data on *process* quality including teacher-child interactions and environment are collected by observers. Data used for this study come directly from the pilot initiative, and data collection procedures follow those outlined by the state's administrative hub.

Teacher education and qualifications. Forms developed by the administrative hub to collect information on teacher education and qualifications were made available to ECE programs who volunteered to participate in Virginia's QRIS. Programs submitted information on the highest degree level of all educational staff members including directors, teachers, and teacher assistants but excluding cooks or family liaisons. Other information related to the subject area of degrees, number of hours of professional development, and information related to other professional development activities such as enrollment in college courses, mentorship, and membership in professional associations.

Information submitted on the self-report forms was reviewed and scored by administrators of Virginia's QRIS, staff at the Virginia Department of Social Services, according to a rubric specifying the number of points earned in this category. Programs could earn up to 40 points toward their total overall QRIS score for this quality component.

Teacher-child ratio and group size. Forms developed by the administrative hub to collect information on teacher-child ratio and group size were made available to ECE programs who volunteered to participate in Virginia's QRIS. Teacher-child ratio and group size information was collected by the programs for every classroom over the time period of one week by recording the number of children and the number of adults present in each classroom for every hour of the day. Programs also submitted information on the birth date of the youngest child in the classroom, and ratios were determined using the age of the youngest child as a guide. These forms were submitted for review and scoring by administrators according to a rubric specifying the number of points earned in this category. Programs could earn up to 30 points toward their total overall QRIS score for this quality component.

Teacher-child interactions. The PreK CLASS (Pianta et al., 2008) and the Toddler CLASS (LaParo et al., 2011) assess classroom interactions between teachers and children and between children. The PreK CLASS is organized along the following 10 dimensions: positive climate, negative climate, teacher sensitivity, regard for student perspectives, behavior management, productivity, instructional learning formats, concept development, quality of feedback, and language modeling. These dimensions create three domains of quality for teacher-child interactions: Emotional Support, Classroom

Organization, and Instructional Support. The Toddler version of CLASS was developed to be appropriate for younger children (15-36 months) and although dimensions and behavior indicators vary slightly, the same three primary domains of interaction are measured in the pilot version of Toddler CLASS used during data collection for the project period.

Reliability and validity of CLASS have been demonstrated in multiple nationwide studies; Chronbach's alpha (internal consistency) of domains ranges from .81 to .94 (Pianta et al., 2008) and scores correlate strongly with other measures of classrooms and teachers (Pianta et al., 2005). PreK CLASS is appropriate for use in classrooms with three- and four-year-old children, and Toddler CLASS is used in classrooms of children under the age of three. Scores are assigned across 10 dimensions of teacher-child interactions, on a scale of 1-7 points. Scores of 1-2 represent low quality, 3-5 represents mid-range quality, and scores of 6-7 are given for high quality teacher-child interactions. In ECE programs where multiple classrooms were observed, scores were averaged across observations and hub administrators applied a rubric to specify the number of points earned toward the program's overall quality rating score. Programs could earn up to 60 points toward their overall QRIS score for this quality component.

Environment. The Early Childhood Environment Rating Scale (ECERS; Harms et al., 2005) and the Infant Toddler Environment Rating Scale (ITERS; Harms et al., 2003) are comprehensive measures of global quality, assessing multiple areas of quality in early childhood environments, including space and furnishings, health and safety, early learning materials, learning formats, and language and interactions. Individual items may

vary according to the age-appropriateness of the group for which the tool is intended, but the tools contain similar subscales for measurement.

The Environment Rating Scales (ERSs) demonstrate good predictive validity (Peisner-Fienberg & Burchinal, 1997; Whitebrook et al., 1990) and internal consistency ($\alpha = .92$; Harms et al., 2005). ECERS is appropriate for use in classroom where the majority of children are 30 months to five years old, and ITERS is used in classrooms where most children are under 30 months old. During an observation using the Environment Rating Scales (ERSs), observers use the appropriate checklist to score programs across items that describe environmental quality. In each item, specific materials or behaviors must be observed, and the item is scored on a scale of 1-7 points. A score of 1 indicates inadequate quality, 3 indicates minimal quality, 5 indicates good quality, and 7 indicates excellent quality. In ECE programs where multiple classrooms are observed, scores were averaged across observations and hub administrators applied a rubric to specify the number of points earned toward the programs overall quality rating score. Programs could earn up to 40 points toward their overall QIRS score for this quality component.

Data Collection Procedures

Data for this study have been collected through Virginia's pilot of their QRIS over the time period between June 2009 and June 2012. Data were collected across the four quality components identified for inclusion in Virginia's QRIS, and were provided as part of an existing database within the Virginia Department of Social Services. Programs submitted self-report forms documenting structural quality prior to their

observation visit, when data on process quality is collected. Programs are re-rated every two years; however, only initial ratings are including in the scope of this study.

Structural quality. Data on *structural* quality such as teacher education and qualifications and ratios and group size were gathered through self-report forms designed for this purpose and submitted for review by the administrators of Virginia's QRIS pilot.

Programs submitted information on the highest degree level of staff members including directors, teachers, and teacher assistants, subject area of staff degrees, number of hours of professional development, and information related to other professional development activities such as enrollment in college courses, mentorship, and membership in professional associations. Information submitted on the self-report forms was reviewed and scored by administrators of Virginia's QRIS, staff at the Virginia Department of Social Services, according to a rubric specifying the number of points earned in this category. Programs could earn up to 40 points toward their total overall score for this quality component.

Teacher-child ratio and group size information was collected by the programs over the time period of one week by recording the number of children and the number of adults present in each classroom for every hour of the day. Programs also submitted information on the birth date of the youngest child in the classroom, and ratios were determined using the age of the youngest child as a guide. These forms were submitted for review and scoring by administrators according to a rubric specifying the number of points earned in this category. Programs could earn up to 30 points toward their total overall score for this quality component.

Process quality. Data regarding the *process* quality of ECE programs were collected through observation according to protocols outlined by the authors of each individual measurement tool, and following the guidelines for observations established by Virginia's QRIS administrators, described below.

A sample of classrooms was observed in each ECE program, and selection of classrooms took place according to the procedure outlined by the administrators. Programs were notified of a three week window in which they may receive an observation visit. The specific date or dates of observation were unannounced, but programs were permitted to select three dates within that window as "blackout" dates on which the observer would not visit the program.

Upon arrival at the ECE program, observers randomly selected one of every three classrooms within the following age ranges: toddler, three-year-olds, and four-year olds (mixed-age groups are considered a separate age range). Only one classroom was observed per visit; therefore, observers visited most programs more than once. According to the structure of Virginia's QRIS, scores on the live observations in selected classrooms represent the quality of the program as a whole. Selected classrooms were observed for a minimum of 4 hours, using the age-appropriate versions of observation tools according to the age of the children enrolled on the day of observation.

PreK and Toddler CLASS. During an observation using the CLASS tools, a 20 minute observation was followed by 10 minutes allotted for scoring. In each classroom, four cycles of observation and scoring were conducted, including at least one cycle for whole-group activities, one cycle for free choice activities, one cycle for meals or transitions, and a fourth cycle of the observer's choice (a second cycle of

meals/transitions was not permitted). According to the tool's authors, PreK CLASS is not conducted outdoors, although Toddler CLASS can be conducted outdoors.

ECERS and ITERS. During an observation using the Environment Rating Scales (ERSs), observers used the appropriate checklist to score programs across items that describe environmental quality. Observers were be present in the classroom for at least 3 hours, but continued longer if necessary. An interview with staff was conducted following the classroom observation to score any items which could not be observed that day.

QRIS observers. Across the state of Virginia, 47 individuals were trained to use the ERS and CLASS observation tools according to QRIS guidelines set by the administrative hub. Raters were trained to use ERS tools by a state anchor who had been trained by the tool's authors; observers must score consistently with the state anchor on 80% of items in order to be qualified to conduct QRIS ratings. Raters also received training through the approved program from the authors of CLASS, and were considered reliable observers upon scoring consistently with 80% of the training program's master codes. To monitor inter-rater consistency, current guidelines developed by the administrative hub require raters to double code and submit scores to the state hub for comparison after every seven ratings, or every 6 months, whichever occurs first.

Calculation of overall quality ratings. Overall ratings were calculated based on the total number of points that a program earned across all four components. State administrators entered information on structural quality into the database from self-report forms, and observers entered information from the observation visit following the final day of observation. Ratings are calculated automatically within the database, by summing

the total number of points earned, and assigning comprehensive ratings according to the following scale: 34-50 earned points is a 1-Star rating, 51-84 earned points is a 2-Star rating, 85-118 earned points is a 3-Star rating, 119-152 earned points is a 4-Star rating, and 153-170 earned points is a 5-Star rating.

Statistical Analysis

Approach. In order to assess whether Virginia's QRIS tiers are depicting meaningfully different levels of quality, it was necessary to first select a method to accurately describe differential levels of quality in participating ECE programs. This study examined the data collected across the four rating components during the pilot phase of Virginia's QRIS to identify qualitatively different subtypes of program quality in ECE programs within Virginia. Following the identification and description of these profiles, correlations were conducted to determine what associations exist between identified profiles and overall quality ratings, to give information about the validity of the comprehensive ratings.

Because the concept of quality is composed of multiple distinct, yet related features, latent profile analysis (LPA) was used to identify the number and type of latent profiles within the sample of participating ECE programs. Latent profile analysis takes a person-centered approach versus a variable centered approach to identifying profiles within data, meaning that this method is uniquely suited to categorize the different subgroups of quality based on the observed data collected through the rating process. In this study, LPA was used to categorize programs based on the type of quality they provide, and each ECE program served as the case in this "person" centered approach.

Profiles were generated from the four components of quality measured during the ratings process outlined by Virginia's QRIS: 1) education qualifications and training, 2) teacher-child interactions, 3) ratios and group size, and 4) environment. Because these four quality components are measured in different ways, it was necessary to standardize scores prior to conducting LPA. Standard scores (Z-scores) were calculated for each component score within each program rating using SPSS version 21.0. Z-scores represent the number of standard deviations a particular observation is with relation to the mean; a positive Z-score indicates an observation above the mean and a negative Z-score indicates an observation below the mean. For the purpose of the state initiative, each component carries a different weight in the child care program's overall quality rating; however, for the purposes of this analysis, they were standardized so that profiles reflect different subtypes of quality as they exist in ECE programs, without the weights applied by Virginia's QRIS rating structure.

It was anticipated that programs' scores across the four components would allow for identification of a certain number of quality profiles within data gathered from Virginia's QRIS during the project period. The distribution of scores demonstrated in the sample through the overall ratings indicated variation in quality, with most ratings falling in the 3-star range and relatively fewer in the 2-star and 4-star category. However, this distribution gave limited information about the *nature* or subtypes of quality in these ECE programs; the structure of the state's QRIS system is such that ratings are assigned based on the total number of points accumulated across the four differently weighted components of quality measured through the system. Any particular program earning a 3-star rating may have earned this score by accumulating points differently across the four

components, meaning that a "3-Star" rating may actually represent any number of different *types* of quality, if types of quality are defined by relative strengths or weaknesses across the four components. Because a particular QRIS rating could theoretically represent multiple subtypes of quality, this study identified profiles of quality and examined the relationship between a program's profile of quality and their overall rating.

Latent profile analysis. Latent profile analysis is a method used for identifying types of related cases within multivariate data. Specifically, LPA is a type of latent variable analysis using continuous observed variables. This same method is referred to as latent class analysis (LCA) when observed variables are categorical (Bartholomew & Knott, 1999). These latent variable analysis methods are used to categorize related cases into classes according to an underlying categorical variable which cannot be observed (Lazarsfeld, 1954). In the current study, each ECE program is a case, and identified classes will describe different subtypes of quality.

Conceptually similar to cluster analysis, this approach is model-based; rather than grouping data into clusters, the analysis estimates the probability of each case's assignment to different classes or profiles. Latent variable analysis offers advantages over traditional cluster analysis; for this study, LPA was conducted using Mplus (Version 6.0; Muthén & Muthén, 1998-2010), which provides statistical fit indices to evaluate and compare models with different numbers of classes. Further, a model-based approach allows for replication with independent datasets (Muthén & Muthén, 2000), which may be particularly advantageous for consideration of revisions or adjustment to state policies.

To address the first research question, LPA was conducted in a series of modeling steps, beginning with the identification of a two-class model and increasing the number of classes until there was no longer improvement in the fit indices (e.g. Chien, et al., 2010; Nylund, Bellmore, Nishina, & Graham, 2007). To assess model fit, the following statistical indicators were used: Akaike Information Criteria (AIC; Akaike, 1987), Bayesian Information Criteria (BIC; Schwartz, 1978), and Adjusted BIC (ABIC; Sclove, 1987).

Pearson's correlation coefficient. To address the second research question, Pearson's correlation coefficient was calculated to examine the linear relationship between a program's profile membership and their comprehensive quality rating. SPSS version 21.0 was used to conduct this analysis.

The following chapter will explain the results of the analyses described above, with findings to address each of the primary research questions examined in this study.

CHAPTER IV

RESULTS

Introduction

This chapter reports findings resulting from the analyses described above to address the following research questions:

1. What patterns or profiles of quality exist among ECE programs participating in Virginia's QRIS?

   a. How are structural and process quality described by the existing patterns or profiles present among participating ECE programs?

2. Is there a relationship between identified profiles and the comprehensive quality ratings assigned by Virginia's QRIS?

   a. What is the relationship between identified profiles and the comprehensive quality ratings assigned by Virginia's QRIS?

Each research question will be addressed in its own section within this chapter.

Descriptive Statistics

This section includes information about the population under study, and the distribution of scores collected during the study period through Virginia's QRIS.

The population under study is early childhood education (ECE) programs in the Commonwealth of Virginia, represented by the sample of programs who have volunteered to participate in the pilot of Virginia's QRIS. Participants have received a total of 358 initial quality ratings during the time period under study. These programs operate in 18 diverse communities across Virginia, in urban, suburban, and rural settings.

Further, participating programs represent a variety of types of care, including small businesses, nonprofit organizations, corporate childcare, religiously affiliated or exempt programs, public pre-kindergartens, Head Start programs, accredited programs, and those accepting childcare subsidies.

For this project, the sample includes all program ratings conducted between June 2009 and June 2012 ($N=$ 358). Star Ratings ranged from 1 Star to 4 Stars ($M = 3.04$, $SD =$ .697). Star Ratings are calculated based on the total number of points earned across the four quality components (maximum 170 points), and total points earned ranged from 49 to 152 ($M = 102.05$, $SD = 21.169$). See Table 1 below for a description of ratings distributions and Figure 1 for the distribution of programs' total score in points.

Table 3
Virginia's QRIS ratings

	1-Star	2-Star	3-Star	4-Star	5-Star
N	1	77	187	93	0

Total Points
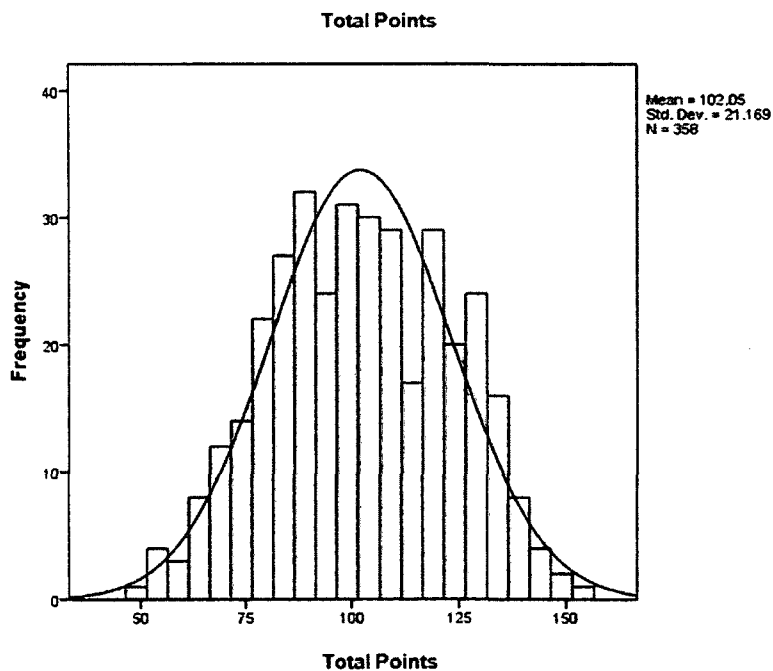
Total Points

Figure 1. Frequency of total points earned by programs.

Descriptive statistics were calculated for each quality component, to describe the program ratings according to their scores across the four data points collected as part of Virginia's QRIS: Teacher education and qualification, Teacher-child interactions, Ratio and group size, and Learning environment. Table 2 below provides a full description of program data in each of these four quality components.

Table 4

Descriptive statistics

		Stars	Total Points	TeacherEdQual	Interactions	RatioGroupSize	Environment
N	Valid	358	358	358	358	358	358
	Missing	0	0	0	0	0	0
Mean		3.04	102.05	22.03	36.65	19.60	23.76
Std. Deviation		0.70	21.17	6.39	9.65	7.18	7.92
Variance		0.49	448.13	40.78	93.04	51.54	62.73
Skewness		-0.10	-0.04	0.24	0.01	-.024	-0.08
Std. Error of Skewness		0.13	0.13	0.13	0.13	0.13	0.13
Kurtosis		-0.78	-0.67	-0.60	-0.50	-0.96	-0.77
Std. Error of Kurtosis		0.26	0.26	0.26	0.26	0.26	0.26
Minimum		1	49	9.00	16.00	6.00	8.00
Maximum		4	152	39.00	60.00	30.00	40.00
Total Possible		5	170	40	60	30	40

Quality Profiles

This section will describe results of analyses conducted to address the first research question. This section includes standardizing the data prior to conducting analyses, the modeling process for latent profile analysis described above including the use of fit indices to select the most sound model, and descriptions of the profiles identified using this analysis.

Standardizing scores. Given that the four quality components each have

different maximums for total points that may be earned within that component, a

standardized Z-score was calculated for programs' scores within each of the four quality

components. Scores were also standardized to help with model convergence in LPA

(Muthén, 2002).

Using these Z-scores, Pearson's correlation coefficient was calculated to

determine any associations between programs' scores in the four quality components.

The findings of this correlation are summarized in Table 3 below.

Table 5

Correlations of standardized scores

	Teacher Education	Interactions	Ratio and Group Size	Environment
Teacher Education	1	.0266[**]	-0.065	0.449[**]
Interactions		1	.210[**]	0.629[**]
Ratio and Group Size			1	0.020
Environment				1

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

Latent profile analysis. To address the first research question, LPA was

conducted in a series of modeling steps, beginning with the identification of a two-profile

model and increasing the number of profiles until there was no longer improvement in the

fit indices (e.g. Chien et al., 2010; Nylund et al., 2007). To assess model fit, the following

statistical indicators were used: Akaike Information Criteria (AIC; Akaike, 1987),

Bayesian Information Criteria (BIC; Schwartz, 1978), Adjusted BIC (ABIC; Sclove,

1987), and the bootstrap likelihood ratio test (BLRT; McLachlan & Peel, 2000).

The first three fit criteria (AIC, BIC, ABIC) indicate how well a model predicts

the observed data for each program. Unexplained variation in the outcome variable (the

latent class variable) increases the value of the fit criteria; therefore, models with lower values on these fit criteria indicate a better fit to the observed data.

Findings indicated that a three profile model is superior to a two profile model, and that a four profile model is superior to a three profile model. When a five profile model was compared to a four profile model, there was not an improvement in fit for the BIC, and only a very small improvement in fit on the AIC and ABIC. As a result, the bootstrap likelihood ratio test (BLRT) was employed as an additional comparison between the four and five profile models. The BLRT was selected because it has been demonstrated to outperform other fit statistics when applied to simulated data set where the correct number of groups is already known (e.g. Henson, Reise, & Kim, 2007).

The BLRT compares an estimated model to a model with one fewer profile, and calculates a p-value to approximate the probability of the data being generated by the model with one fewer profile. In using this test, a low p-value indicates that the model with the lower number of profiles can be rejected in favor of the model under examination. The BLRT was used to confirm that a four profile model is preferable to a three and a two profile model. When the five profile model was compared with a four profile model, the BLRT did not yield a lower p-value, indicating that the five profile model did not offer a significant improvement in fit, compared to the four profile model. Therefore, the four profile model was selected as the most parsimonious description of programs in Virginia's QRIS pilot. Table 4 below gives the fit indices for model comparison.

Table 6

Fit indices for model comparison

	AIC	BIC	ABIC	Log likelihood	BLRT p-value
2 profile model	3856	3907	3866	--	--
3 profile model	3816	3886	3826	-1915.49	0.0000
4 profile model	3786	3875	3803	-1890.46	0.0000
5 profile model	3780	3889	3800	-1870.37	0.0500

Description of profiles. Each of the four profiles in the model describe a different type of quality within the participant programs of Virginia's QRIS, and Table 5 gives estimated Z-score means for each variable in the analysis and the size of each profile within this model.

Table 7

Profile estimated means and membership

Quality Component	Four Profile Model			
	P1	*P2*	*P3*	*P4*
Teacher Education	-0.68	-0.45	0.64	1.09
Interactions	-0.97	0.07	0.98	-0.42
Ratio and Group Size	-0.08	0.10	0.36	-0.92
Environment	-1.23	-0.03	0.97	0.36
N (profile membership)	96	117	105	40
Total *N* = 358				

In Figure 2 below, profiles of programs are graphically described using the estimated means from Table 5 above. The *y*-axis shows the mean of the standardized Z-scores.
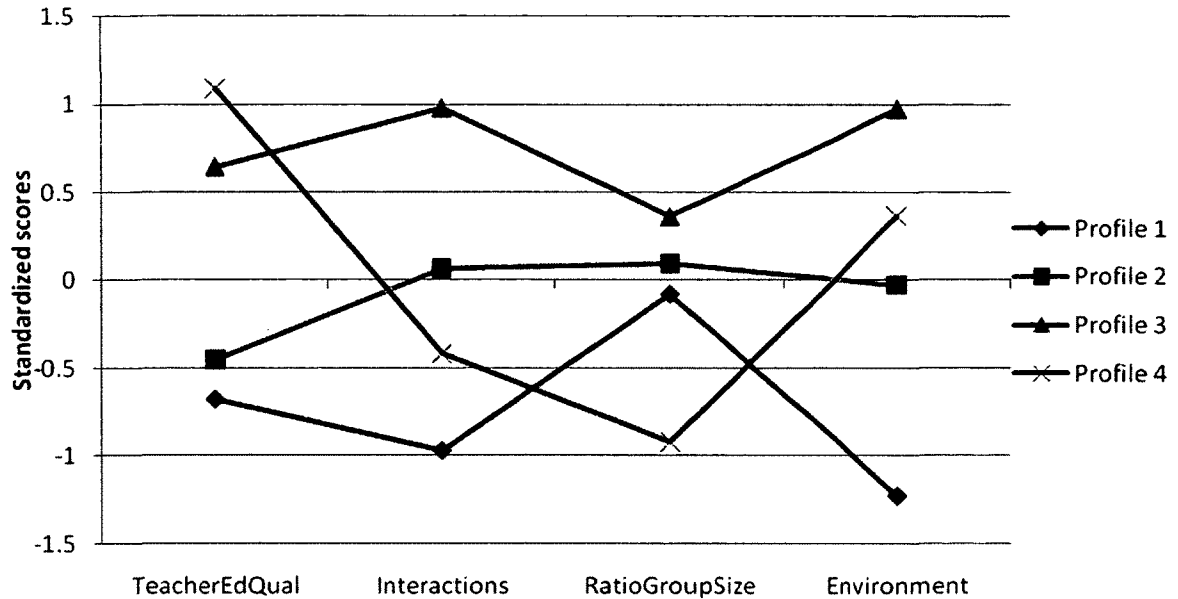
Figure 2. Four profile model

In the first profile, *Basic Quality* (*n*=96), programs score below the mean in each

of the quality components, including both structural and process quality features. The

Basic Quality profile has the lowest scores of all profiles for three of the four components

(Teacher Education, Interactions, and Environment). This profile represents

approximately 26.8% of the programs participating in Virginia's QRIS pilot.

Table 8
Profile 1: Basic quality (n=96)

Quality component	Minimum Z-score	Maximum Z-score	*M (SD)*
Teacher education	-2.04	1.14	-0.66 (0.72)
Interactions	-2.14	0.76	-0.98 (0.66)
Ratio and group size	-1.89	1.45	-0.11 (0.94)
Environment	-1.99	-0.22	-1.27 (0.42)

The second profile, *Mean Quality* (n=117) is the largest profile, representing

32.6% of participating programs. Although scores in Teacher Education are lower than

average, programs in the *Mean Quality* profile score near the mean in three of the four

quality components (Interactions, Ratio and Group Size, and Environment), including

both structural and process quality features.

Table 9
Profile 2: Mean quality (n=117)

Quality component	Minimum Z-score	Maximum Z-score	*M (SD)*
Teacher education	-2.04	0.93	-0.50 (0.68)
Interactions	-1.73	2.00	0.08 (0.66)
Ratio and group size	-1.89	1.45	0.08 (0.95)
Environment	-0.98	1.29	-0.00 (0.49)

In the third profile, *Global Quality* (n=105), programs score above the mean in

each of the quality components, including both structural and process quality features.

The *Global Quality* profile has the highest scores of all profiles for three of the four

components (Interactions, Ratio and Group Size, and Environment). This profile

represents approximately 29.3% of the programs participating in Virginia's QRIS pilot.

Table 10
Profile 3: Global quality (n=105)

Quality component	Minimum Z-score	Maximum Z-score	*M (SD)*
Teacher education	-1.10	2.34	0.69 (0.73)
Interactions	-0.48	2.42	1.00 (0.63)
Ratio and group size	-1.89	1.45	0.41 (0.88)
Environment	-0.48	2.05	1.00 (0.56)

The fourth profile, *Variable Quality* (*n*=40) is the smallest profile, representing

11.2% of participating programs. Programs in this profile had the highest mean scores in

Teacher Education and the lowest mean scores in Ratio and Group Size (both structural

features). Scores in process quality varied also, with Interactions scores as a relative

weakness and Environment scores as a relative strength.

Table 11

Profile 4: Variable quality (n=40)

Quality component	Minimum Z-score	Maximum Z-score	*M (SD)*
Teacher education	-0.31	2.66	1.25 (0.65)
Interactions	-2.14	0.76	-0.51 (0.66)
Ratio and group size	-1.89	0.95	-1.06 (0.77)
Environment	-0.48	1.30	0.42 (0.57)

Star Ratings

This section will describe results of analyses conducted to address the second

research question. This section includes examination of the relationship between the

profiles identified above and the comprehensive Star Ratings assigned by Virginia's

QRIS.

Pearson's correlation coefficient. To address the second research question,

Pearson's correlation coefficient was calculated to examine the linear relationship

between a program's profile membership and their comprehensive quality rating.

Correlations were also calculated between profiles to determine the degree to which the

profiles represented distinct constructs of quality.

Correlations between different quality profiles range from small ($r(356) = .215$, p

< .01) to moderate ($r(356) = .449$, $p < .01$), indicating that profiles are successfully

identifying distinct constructs in the participating programs. Table 10 provides the results

of the correlations between overall Star Ratings and profile membership.

Table 12

Pearson's correlation for ratings and profile

	Overall Stars	Overall Score	Profile 1	Profile 2	Profile 3	Profile 4
Overall Stars	1	0.909**	-0.686**	-0.065	0.748**	-0.020
Overall Score		1	-0.703**	-0.061	0.760**	-0.019
Profile 1			1	-0.422**	-0.390**	-0.215**
Profile 2				1	-0.449**	-0.247**
Profile 3					1	-0.228**
Profile 4						1

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

Results indicate that for *Basic Quality* and *Global Quality* profiles, there is a

strong correlation between a program's comprehensive Star Rating and profile

categorization. *Basic Quality* profile membership is related to a lower Star Rating, and

Global Quality profile membership is related to a higher Star Rating. For other profiles

(*Mean Quality* and *Variable Quality*), there is not a strong relationship between the

comprehensive Star Rating assigned by Virginia's QRIS and the probability of being

categorized into either profile.

The next chapter is devoted to discussing the findings outlined above, and

situating the results of the current study in the field of research in QRIS. The chapter

includes a brief overview of the study including major findings, as well as an

interpretation of results in terms of the research questions. The chapter also addresses a

discussion of some of the limitations of the study and recommendations for future

research.

CHAPTER V

DISCUSSION

Introduction

The purpose of this study was to yield information regarding the validity of Virginia's QRIS levels by first identifying quality profiles using data gathered as part of programs' participation in the QRIS, and second by examining the relationship between those quality profiles and the comprehensive Star Ratings assigned by the pilot initiative. This chapter begins with a brief overview of the study, a summary of major findings and interpretation these findings. The chapter concludes with a discussion of some of the limitations of the study and recommendations for future research.

Study Overview

Each year, increasing numbers of children participate in some type of non-parental care prior to kindergarten. In 2011, 42% of American four-year-olds attended some type of public preschool programs (such as state pre-kindergarten or Head Start), with an additional 32% in other types of center-based care, and another 19% attending some type non-relative family child care (Barnett et al., 2012)

The level of early childhood care and education that children receive in these settings makes a difference to their school readiness outcomes in the academic and social-emotional domains, in addition to impacting adult life outcomes. To address the need to monitor investment in early childhood and assure that outcomes of such programs are positive, many publicly funded programs have launched initiatives or mandates to outline what high quality care would look like in these settings. Two examples of these

examples of these initiatives are the program quality standards released by the National

Institute for Early Education (NIEER) for public prekindergarten programs, and

improvement initiatives from the Office of Head Start (OHS) to improve quality in Head

Start grantees.

However, a large number of children attend non-public preschool programs (51%

of four-year-olds), and these settings represent a variety of program types, including

private businesses, non-profit organizations, family child care homes, and programs

exempt from licensure for affiliation with religious organizations. This variety of settings

reflects an uncoordinated early childhood system that presents a challenge for improving

quality across the board for young children.

QRIS initiatives were launched by states in an effort to address an

acknowledgement that although child care licensing assured that basic health and safety

standards were being met in early childhood settings, there was little information

available about the quality of these programs beyond those basics. Further, although

private accreditation programs are an alternative for programs wishing to be designated at

the highest quality levels, few programs had the capacity or resources to achieve these

high thresholds (Westervelt et al., 2008).

Nevertheless, QRIS initiatives are situated in this landscape of quality alongside

increased standards for public programs, efforts to improve and strengthen licensing

requirements, and private accreditation programs. For this reason, most states build their

QRIS programs on a foundation of licensing—that is, programs must be in good standing

with the regulatory authority before being accepted for QRIS (Tout et al., 2010).

Additionally, some states elect to include accreditation as a pathway for programs to

reach the highest tiers of QRIS (Tout et al., 2010), recognizing that accredited programs have achieved very high thresholds of quality that often place them in the upper tiers of such a system.

QRIS initiatives have gained increasing momentum across the country. In a review of state QRISs conducted in 2010, 26 states had implemented a QRIS initiative. Recently, an updated map of QRIS participation was released by the QRIS National Learning Network (2013), with 38 states having launched a QRIS statewide, two states with a regional QRIS, two states with pilot QRIS, and 11 states planning for a QRIS.
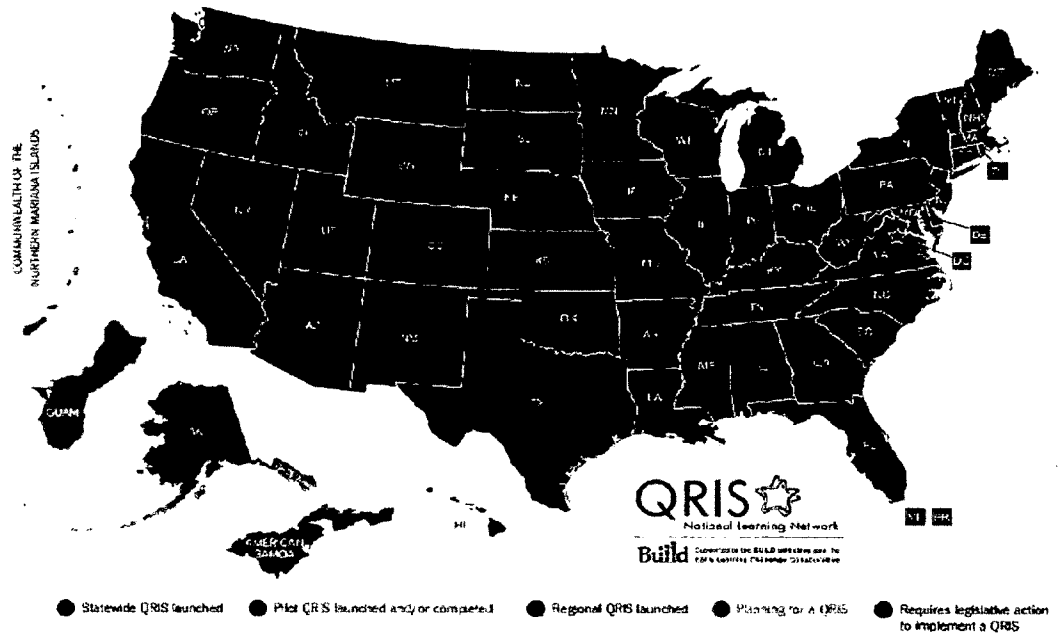
Figure 3. 2013 QRIS map (QRIS NLN, 2013)

Rapid growth in the expansion of QRIS across the country is supported by funding initiatives from the federal government such as the Race to the Top- Early Learning Challenge Grant, which requires states to operate a QRIS to maximize grant awards (U.S. Department of Education & U.S. Department of Health and Human Services, 2011) and the Child Care Development Fund requirement that a portion of

funds be dedicated to quality improvement (U.S. Department of Health and Human Services, 2013). Increasingly, states are encouraged to develop QRISs in order to access as much funding as possible for their early childhood systems.

The evidence base for developing these systems is still small, and each state develops and operates its QRIS independently. Thus, each state's system necessitates individual examination of the degree to which the QRIS is fulfilling its intended purpose within that state. Due to the reporting requirements on some of the QRIS-related funding streams, it is likely that the evidence base for these systems and their implementation will increase.

Validation of QRIS frameworks includes research activities that examine whether the structure of a QRIS meaningfully differentiates levels of quality. Validating a QRIS system is a process of justifying the decisions about how tiers are delineated by demonstrating that they link to another method of differentiating quality. Zellman and Fiene (2012) emphasize that validation is a multi-step process to consider the accuracy and meaningfulness of ratings. The current study aims to contribute information regarding the validity of Virginia's QRIS system by examining how quality profiles among data on four quality components relate to the comprehensive Star Ratings assigned by the statewide initiative. In addition, the current study contributes to the body of research on QRISs nationally, as an example of validation research that is necessary to build the evidence base for implementing such initiatives.

Using quality data collected from the pilot programs participating in Virginia's QRIS, latent profile analysis was used to identify distinct quality profiles within the sample of programs participating in Virginia's QRIS pilot. Correlations were then

conducted to yield information about how these distinct subtypes of quality relate to the comprehensive Star Ratings.

Summary of Findings

Data analysis revealed a four-profile model of quality profiles as the most parsimonious description of the quality profiles within the sample programs. Four distinct quality profiles were identified using LPA: *Basic quality, Mean quality, Global quality,* and *Variable quality.*

Basic quality describes a profile of programs that scored below the mean in each of the quality components in Virginia's QRIS. The *Mean quality* profile is characterized by quality components that cluster around the mean, and were neither distinctly high nor low quality, compared with other profiles. Programs in the third profile, *Global quality,* had above average scores on each of the four quality components, and achieved higher levels of overall global quality (both structure and process quality) than any other profile identified in the analysis. *Variable quality* programs are a small segment of the sample with a unique pattern of relative strengths and weaknesses compared to the three other profiles.

Strong, significant relationships between profile membership and Star Rating were identified for two profiles (*Basic quality* and *Global quality*), and no significant relationships were identified for the other two profiles (*Mean quality* and *Variable quality*). A negative, strong relationship was identified between the *Basic quality* profile and Star Rating, indicating that membership in this profile was associated with a fewer stars in the Star Rating. In contrast, a positive, strong relationship was identified between the *Global quality* profile and Star Rating, indicating that membership in this profile was

associated with more stars in the Star Rating. This finding is in alignment with the hypothesis for this study; some profiles are more strongly associated with higher Star Ratings compared with other profiles, indicating that the framework itself is not discerning levels of quality in a consistent way when related to quality profiles. This finding yields uneven validity evidence for Virginia's QRIS.

Interpreting Findings

To participate in Virginia's QRIS, all programs must be in good standing with basic health and safety standards according to the appropriate regulating authority (e.g., childcare licensing, Virginia Preschool Initiative standards, or Head Start regulations) (VSQI, 2009). Although all programs within the *Basic quality* profiles are meeting minimum standards for basic health and safety, the programs in this quality profile are generally scoring below average compared to other profiles on most quality components included in Virginia's QRIS. Low quality in the area of Teacher education and Environment may indicate low levels of financial resources in *Basic quality* programs, given that these components require funding to hire qualified staff and purchase materials and resources for classrooms.

When comparing *Mean quality* programs with *Basic quality* programs, there were not major differences in the structural quality components (Teacher education and Ratio and group size). *Mean quality* programs did have higher quality than *Basic* in each component, but particularly in Interactions and Environment. Because Interactions and Environment are the two components that are observed during an on-site observation measuring process quality, this could indicate that programs in the *Mean quality* profile have a relative strength in process quality compared to structural quality, or that they

were relatively better prepared for their on-site observation compared with programs whose scores indicated *Basic quality* profile membership. Another possible interpretation is that process quality may be more malleable, as measured in Virginia's QRIS, than structural quality in the field. The nature of QRIS frameworks is such that comprehensive quality ratings are constructed out of qualitatively different components or data points collected in the field. For this reason, it is difficult to discern the relative effects that each element of quality could have, either on the overall rating, or on its impact for children attending that program.

Practically speaking, quality components that can be improved through teacher training or professional development offer opportunities for improvement in ways that structural quality such as degree levels of teachers or ratios may not. Changes to structural quality may be more feasible through regulation or legislation of requirements for staff or maximums on group size for preschool classrooms, and therefore, it may be worth considering whether their inclusion in QRISs is valuable for incentivizing quality improvement.

Although programs in the *Global quality* profile scored higher than *Basic* and *Mean* in each component, Teacher education remains a relative weakness compared with other components in the *Global quality* profile. This may indicate that even when programs are able to achieve above average quality overall, this component remains challenging to improve. Some possible explanations for this could be that staffing represents a significant cost to programs. Or, this finding could be reflective of the points-based framework which reserves the highest number of points in this component for levels of Teacher education that are especially challenging, such as Masters degrees.

If very few programs are achieving the highest number of points in this quality component, this may be an area where a re-evaluation of the assignment of points is warranted.

Although it was the smallest profile, *Variable quality* had the most unique composition when compared with the other three profiles. *Variable quality* programs have the highest scores in Teacher education, and nearly the highest scores in Environment. In the *Basic quality* profile, these two components were weaknesses, indicating a potential lack of resources to bolster scores in these areas. In contrast, these components are strengths for *Variable quality* programs, possibly indicating that programs in this category are well funded by comparison.

Interestingly, the *Variable quality* profile scored poorly in Ratio and group size, with scores well below the mean. This patterning is especially unique since each of the three other profiles' scores cluster around the mean for this quality component. The fact that Ratio and group size scores do not vary as widely between the first three profiles is possibly due to economic constraints on programs to enroll as many children as permitted by their regulating authorities. Programs may not be able to afford to improve their ratios when it affects the bottom line of their businesses. When all programs (regardless of other quality markers) are maintaining ratios in compliance with, for example, the Department of Social Services, the Ratio and group size component is less able to differentiate among different subtypes of quality within the sample of programs in this study.

It is possible that the programs categorized by the *Variable quality* profile are regulated by a system with different ratio requirements, or that these programs are not

subject to the same economic constraints as other programs. The particularly high Teacher education and Environment quality scores indicate that these programs have adequate funding to hire qualified staff and resource classrooms generously. Although data are not available to examine whether these programs are classified as publicly funded programs, it is possible that publicly funded programs with high teacher qualification requirements and higher enrollment, would exhibit the same patterning of quality components as those programs in the *Variable quality* profile. Public programs would not suffer from consolidation of classrooms and other issues impacting Ratio and group size because even in difficult economic times, programs which are free of charge remain full.

Another possible interpretation of this profile is that these programs are engaging in a different instructional approach that prioritizes child-led exploration (rather than teacher-child interactions) with an emphasis on enriched environments, such as Montessori or Reggio inspired programs. However, without additional information on program type, it is not possible to determine the exact nature of this profile.

For programs in the *Basic quality* and *Global quality* profiles, the comprehensive rating system appears to be accurately differentiating between these two types of quality. However, for programs categorized as *Mean quality* or *Variable quality*, no significant relationship was detected between Star Rating and profile membership. When interpreting these differences, it is useful to consider the points-structure framework of Virginia's QRIS. Star Ratings are assigned by summing the total number of points earned by a program across all four of the quality components measured by the system. Specific ratings are assigned according to predetermined cut-off scores between tiers.

Therefore, in order to earn the highest Star Ratings, a program must earn a substantial number of points across most quality components; if a program were to score particularly poorly in one or more quality component, it would be mathematically difficult to accumulate enough points to earn a high Star Rating. The same logic applies to programs with the lowest Star Ratings; in order to earn a low Star Rating, a program would have to score relatively poorly across most of the quality components.

For this reason, it is reasonable to expect that *Basic quality* is associated with lower Star ratings because this profile is characterized by lower scores across the four quality components. By the same token, *Global quality* is characterized by higher scores in the four quality components, which would also be associated with the potential to earn higher Star Ratings. For programs with high overall or low overall quality profiles, the comprehensive Star Rating is a valid means of differentiating quality levels. However, validation of QRIS tiers imply that the system is functioning adequately to differentiate quality at every rating level, which was not the finding of the current study.

In order for a QRIS to function in its role to accurately measure and communicate meaningfully different levels of quality in early childhood programs, it is essential that the tiers of the QRIS depict different levels or types of quality in a way that can be demonstrated through other means. The findings of the current study do not indicate that the tiers of Virginia's QRIS are accurately depicting different levels of quality, when quality is defined by the profiles or subtypes revealed in this analysis. However, this does not guarantee that there are not other means of conceptualizing quality that may be accurately demonstrated through comparison with comprehensive Star Ratings, pointing to the need for a multi-faceted approach to validation research.

In a points-based QRIS, programs with middle-range Star Ratings may earn these ' ratings in a variety of ways. Programs can earn middle-range Star Ratings either by scoring moderately well in each of the quality components, or they may earn the same number of points overall by having a balance of strengths and weaknesses across the quality components. There are many combinations of relative strengths and weaknesses that can result in the same middle-range Star Rating.

This may explain why profiles with moderate scores (*Mean quality*) or combinations of scores (*Variable quality*) do not show strong associations with Star Ratings. Quality profiles represent patterns of strengths and weaknesses, but not particular combinations of points. In terms of validation of the QRIS tiers, Virginia's QRIS does not adequately differentiate quality when quality is defined by the specific profiles identified in the current study.

Because validation activities vary widely and there is no prescribed methodology for validating a QRIS (Tout & Starr, 2013; Zellman & Fiene, 2012), it would be advisable to engage in other validation efforts to gather additional information about how the tiers of Virginia's QRIS relate to objective information about how quality differs in centers with different Star Ratings. If the results of additional investigations yield similar results, stakeholders may consider revisions to Virginia's framework to strengthen its ability to meaningfully differentiate between tiers.

Two specific revisions may be worth considering. First, the structure of a points-based system may make it more difficult to relate Star Ratings to other quality data. As discussed above, it is possible to earn a middle-range Star Rating through a number of different combinations of strengths and weaknesses, meaning that it is unlikely that all 3-

Star ratings, for example, are a particular type of quality. Further, a final validation activity is to compare overall quality ratings with the developmental outcomes of children in differently rated centers. Thinking forward to this stage of validation, it seems unlikely that children in 3-Star centers would have different developmental outcomes compared with children in lower or higher rated centers, given that these middle-range ratings can be achieved through various combinations of strengths and weaknesses in quality component areas.

Second, the inclusion of quality components which are constrained by economic factors may distort the meaningfulness of ratings. When all programs are constrained economically to have similar ratios and struggle to raise the qualifications of their staff, these factors serve less purpose in differentiating quality levels. It is possible that a more streamlined QRIS with fewer components may be easier to validate. Importantly, states include components in their QRIS frameworks for a variety of reasons (e.g. representing the values of the leadership, maintaining partnerships with stakeholders, incentivizing program improvement), and therefore considering addition or deletion of components must be a thoughtful process (Wesley & Buysse, 2010).

Limitations

This study's findings are limited in three main ways: quality of data, generalizability, and limitations to the ability of findings to validate the structure of Virginia's QRIS.

Data. Two of the four components of quality measured by Virginia's QRIS are collected through self-report forms generated by the administrative hub for use in calculating quality ratings to be assigned to participating programs (VSQI, 2009). These

forms were not developed for research purposes, and given the high-stakes nature of published QRIS ratings, programs may have misrepresented information about their qualifications or ratios in order to receive a higher rating. The administrators of QRIS engage in periodic spot-checking of this information, but these efforts are not well documented. Observation data are limited because information about the enforcement of inter-rater consistency guidelines is not available for the period of June 2009 to June 2012 (VSQI, 2012).

Generalizability. The ability to replicate the analyses in this study holds unique potential to extend work from this project to inform potential changes to a Virginia's QRIS, or examining questions about how the landscape of class membership might change over time as a state continues to invest in quality improvement efforts. Findings from this study will not be generalizable to QRIS programs outside of Virginia, and will yield information only about programs participating in the pilot of Virginia's QRIS. Participants in the pilot of Virginia's QRIS are not necessarily representative of ECE programs throughout the state, or in other states. The pilot phase of this initiative was voluntary, meaning that it is possible that programs volunteering to participate in the pilot share qualities that make them different from other programs throughout the state.

Furthermore, the pilot of Virginia's QRIS was only available in 18 volunteer communities across the state, not selected to be representative of the state population, and it is possible that communities who volunteered for participation in the pilot are also communities with more supportive early childhood systems. Future studies will benefit from Virginia's current efforts to expand the QRIS initiative throughout all areas of the state. In the most recent request for proposals for organizations to implement Virginia's

QRIS locally, the state required that all regions of the state be served through this initiative (VSQI, 2013).

Validation. The findings of this study have the potential to contribute to a discussion regarding validating the structure of Virginia's QRIS, and validation approaches for other states QRIS intiatives. However, individual validation activities can only provide partial information about how well a structure is functioning in its goal to accurately describe levels of quality in ECE programs (Zellman & Fiene, 2012). Validation is a category of research activities that encompasses multiple approaches, and validation is a multi-step process such that no single validation activity will allow a state to draw the conclusion that a QRIS is either valid or invalid (Tout & Starr, 2013). Instead, the present study seeks to provide information about how the tiers of Virginia's QRIS relate with profiles of quality identified within the ECE programs participating during the pilot phase.

Future Research

Virginia's QRIS. There are several ways that findings from this study can be extended and deepened to learn more about Virginia's QRIS. In the future, it may be possible for Virginia to collect and store raw data from the four quality components, so that these data may be analyzed directly rather than data which have been converted to points within Virginia's QRIS framework. Regarding additional data collection that would further this research, it would be interesting to collect quality component data from non-QRIS programs, to determine whether quality profiles identified in this study are present in the population of programs who did not volunteer for this pilot initiative.

Future investigations may examine whether profiles are maintained when using additional quality components that are not included in Virginia's rating system, such as salary for teachers, parent engagement, or curriculum use. Information on program type, geo-coding, size, enrollment percentage, and cost to parents would also facilitate interpretation of the quality profiles, although these analyses were outside the scope of this study.

If Virginia continues to conduct validation activities to examine the accuracy of the points-based system in place, logical next steps would include collection of an independent quality measure in a representative sample of programs. This approach would allow for investigation of how well QRIS ratings relate to a different measure of global quality, such as the Assessment Profile (Abbott-Shim & Sibley, 1989) or the Observational Record of the Caregiving Environment (ORCE; NICHD Early Child Care Research Network, 1996). Conducting analyses to examine how QRIS ratings relate to an external measure of global quality would contribute an important piece of information to the discussion of validation; however, it is important to select a tool for this purpose that is in alignment with the priorities of Virginia's QRIS. For example, it would be unproductive to select an external quality measure with a heavy emphasis on literacy, since this was not a priority in the development of Virginia's QRIS, and therefore would be unlikely to relate strongly to the comprehensive ratings.

A more resource intensive validation activity would be carefully selecting and collecting child outcome data in QRIS programs to discern whether QRIS levels translate to meaningful differences in children's learning and development. Of particular

importance may be studies which examine how much and how quickly children grow in programs that receive different Star Ratings.

If Virginia considers making revisions to better align with other states, it may be helpful to statistically model other states' QRIS frameworks using Virginia data, to observe the effects on ratings or distributions of ratings under another QRIS framework. This may be particularly helpful in considering a building-blocks or combination framework, rather than a points-based structure.

Validation. Validation studies of QRIS to date have been limited. However, this is expected to change as states with Race to the Top- Early Learning Challenge grants (U.S. Department of Education, & U.S. Department of Health and Human Services, 2011) are required to meet the validation study component of this funding. One of the first validation studies focused on Colorado's QRIS (Zellman et al., 2008), and findings from this research highlighted the challenges for states to demonstrate that the tiers of a QRIS relate meaningfully to other measurements of childcare quality. Some less formal validation activities such as examining the validity of QRIS components are likely underway in many states, and states such as Massachusetts (Schilder, Young, Anastasopoulous, Kimura, & Rivera, 2011) and Minnesota (Minnesota DHS, 2011); have formalized this process through more recently published reports.

In Virginia, informal examinations of the validity of components have been conducted, in addition to a more formal validation study conducted by Sabol and colleagues (2011) examining the link between QRIS ratings and children's development in the Virginia Preschool Initiative. However, given that validation studies are multi-

faceted and multi-phased, additional research studies will be needed to generate a comprehensive picture of the validity of Virginia's QRIS.

Findings from the current study are relevant for policymakers in Virginia, but restrictive budget climates in state governments make low-cost research activities particularly valuable. A practical direction for research in QRIS is the area of maximizing existing data to inform decision-making. In part, these studies could make use of administrative data such as child care subsidy data (e.g. attendance) or mandatory child outcome measures through a state's prekindergarten program. A range of data are available in various state QRIS databases, depending both on the quality components selected for inclusion in the state's rating system, in addition to data that are administratively gathered as part of programs' participation. States may already be collecting information regarding improvement activities that take place in the QRIS, including mentoring hours and strategies, and the expenditure of incentive funds provided as part of participation.

Analytic strategies that can make the most of existing data without having to devote resources to collection of new data can help to ensure that a state's QRIS is meeting the goals for which it was established, without continuing to invest additional resources into a system that may not be performing adequately. In the course of a state's QRIS development, revisions may be considered as state administrators receive feedback from programs and other stakeholders regarding successes and challenges of implementation. And, as states are considering revisions, using existing quality data to model the impacts of adjustments to the rating system can help ensure informed decision-making and generate solutions to anticipated concerns.

Conclusion

The findings from this study indicate a need for further investigation into the relationship between comprehensive Star Ratings assigned by Virginia's QRIS and other means of conceptualizing quality, as stated in the previous section. However, the findings may still inform a forthcoming round of revisions planned by administrators and stakeholders for the rating system.

In particular, because Ratio and group size data did not demonstrate major differences for three of the four profiles representing nearly 90% of participating programs, it is possible that Ratio and group size does not differ widely among different types of quality in child care programs participating in Virginia's pilot. Should Virginia approach the revisions process with a goal of parsimony and eliminating components that do not actively differentiate between types of quality, Ratio and group size may be a component that is considered for deletion.

Given that the points-based structure of Virginia's QRIS may make it more difficult to differentiate quality in programs with middle-range quality ratings, it is possible that consideration of another framework could increase the validity of tiers. One advantage of the points-based structure is that all four quality components are collected for every participating program. In building-blocks and combination structures, this is not always this case, because each tier adds onto and builds on the quality components of the previous tier. Because of this data collection, Virginia is well-positioned to engage in research activities that model potential new structures using the data already collected through pilot phases.

Another consideration for adjusting the framework of Virginia's QRIS is that points-based QRISs can be more expensive to implement because of the intensive data collection. This consideration must be balanced by the recognition that if Virginia adopts a new structure, it will not have access to the rich database of complete quality component information that it currently maintains.

The current study also has implications for the quality improvement phase of QRIS. The majority of profiles, representing the majority of programs, demonstrated Teacher education as a relative weakness. As suggested above, it is possible that the high cost of continued higher education could be related to this pattern. Allocating funding for programs to increase the qualifications of their staff may be an opportunity to support programs' quality improvement in a critical area.

Analyses conducted for this study revealed several distinct types of quality within the sample of child care programs volunteering to participate in Virginia's pilot QRIS. These quality profiles describe the various patterns that exist in the quality component data collected through the rating initiative. By examining the relationship between distinct types of quality and the comprehensive Star Ratings assigned by Virginia's QRIS, the findings of this study are important for Virginia's future decision-making regarding the structure of its rating system. There were strong relationships between some of the quality profiles identified in the analysis, but other profiles did not demonstrate a relationship with the Star Ratings. This study raises questions regarding why certain profiles relate more strongly to Star Ratings than others, and points to the need for further investigation into the distinctions that are made between tiers of Virginia's QRIS.

REFERENCES

Abbott-Shim, M., & Sibley, A. (1987). *Assessment Profile*. Atlanta: Quality Assist.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*, 716-723.

Antle, B.F., Frey, A., Barbee, A., Frey, S., Grisham-Brown, J., & Cox, M. (2008). Child care subsidy and program quality revisited. *Early Education and Development, 19*(4), 560-573.

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology, 10*, 541-552.

Barnard, W., Smith, W. E., Fiene, R., & Swanson, K. (2006). *Evaluation of Pennsylvania's Keystone STARS quality rating system in child care settings.* Pittsburgh, PA: University of Pittsburgh School of Education, Office of Child Development.

Barnett, W.S., Carolan, M.E., Fitzgerald, J., & Squires, J.H. (2012). *The state of preschool 2012: State preschool yearbook.* New Brunswick, NJ: National Institute for Early Education Research.

Bartholomew, D.J., & Knott, M. (1999). *Latent variable models and factor analysis.* London, UK: Arnold.

Burchinal, M.R., & Cryer, D. (2003). Diversity, child care quality, and developmental outcomes. *Early Childhood Research Quarterly, 18*, 401-426.

Burchinal, M., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes? In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.),

Quality Measurement in Early Childhood Settings (pp. 11-31). Baltimore: Brookes.

Burchinal, M., Nelson, L., Carlson, M., & Brooks-Gunn, J. (2008). Neighborhood characteristics and child care type and quality. *Early Education and Development, 19*(5), 702-725.

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly, 25,* 166-176.

Campbell, F.A., & Ramey, C.T. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families. *Child Development, 65,* 684-698.

Cassidy, D.J., Hestenes, L.L., Hansen, J.K., Hegde, A., Shim, J., Hestenes, S. (2005). Revisiting the two faces of child care quality: Structure and process. *Early Education and Development, 16*(4), 504-520.

Cheatham, J., Pope, B., & Myers, G. (2005). *Evaluating quality in state child care licensing: The Tennessee report card and star-quality child care program.* The University of Tennessee College of Social Work Office of Research and Public Service.

Chien, N., Howes, C., Burchinal, M., Pianta, R., Ritchie, S., Bryant, D., Clifford, R., Early, D. M., & Barbarin, O. (2010). Children's classroom engagement and school readiness gains in prekindergarten. *Child Development, 81,* 1534-1549.

Curby, T.W., LoCasale-Crouch, J., Konold, T.R., Pianta, R.C., Howes, C., et al. (2009). The relations of observed pre-k classroom quality profiles to children's achievement and social competence. *Early Education and Development, 20*(2), 346-372.

Currie, J. (2000). *Early childhood intervention programs: What do we know?* Northwestern University/University of Chicago Joint Center for Poverty Research.

Dominguez, X., Vitello, V.E., Maier, M.F., & Greenfield, D.B. (2010). A longitudinal examination of young children's learning behavior: Child-level and classroom-level predictors of change throughout the preschool year. *School Psychology Review, 39*(1), 29-47.

Downer, J. (2013). Applying lessons learned from evaluations of model early care and education programs to preparation for effective implementation at scale. In T. Halle, A. Metz, & I. Martinez-Beck. (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 157-169). Baltimore, MD: Brookes.

Downer, J.T., Booren, L.M., Lima, O.K., Luckner, A.E., & Pianta, R.C. (2010). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. *Early Childhood Research Quarterly, 25*, 1-16.

Dowsett, C.J., Huston, A.C., Imes, A.E., & Gennetian, L. (2008). Structural and process features in three types of child care for children from high and low income families. *Early Childhood Research Quarterly, 23*, 69-93.

Halle, T., Martinez-Beck, I., Forry, N.D., & McSwiggan, M. (2011). Setting the context for a discussion of quality measures. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality Measurement in Early Childhood Settings* (pp. 3-10). Baltimore: Brookes.

Hamre, B.K., & Pianta, R.C. (2005). Can instructional and emotional support in the first grade classroom make a difference for children at risk of school failure? *Child Development, 76,* 949–967.

Harms, T., Clifford, R.M., Cryer, D.R. (2005). *The Early Childhood Environmental Rating Scale- Revised.* New York: Teachers College Press.

Harms, T., Cryer, D.R., & Clifford, R.M. (2003). *Infant/Toddler Environmental Rating Scale- Revised.* New York: Teachers College Press.

Helburn, S.W. (1995). *Cost, quality and child outcomes in child care centers: Technical report.* Denver: University of Colorado, Department of Economics, Center for Research in Economic and Social Policy.

Helburn, S.W., & Bergmann, B. (2002). *America's child care problem: The way out.* New York: Palgrave.

Henson, J.M., Reise, S.P., & Kim, K.H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling: A Multidisciplinary Journal, 14,* 202-226.

Howes C., Burchinal M., Pianta R., Bryant D., Early D., Clifford R., Barbarin O. (2008). Ready to learn? Children's pre-academic achievement in prekindergarten programs. *Early Childhood Research Quarterly, 23,* 27–50.

Howes, C., Phillips, D. A., & Whitebook, M. (1992). Thresholds of quality: Implications

for the social development of children in center-based child care. *Child*

Development, 63, 449-460.

Invernizzi, M. A., Sullivan, A., Meier, J. D., & Swank, L. (2004). *Phonological*

awareness literacy screening for preschool: Teacher's manual. Charlottesville,

VA: University of Virginia.

Johansen, A. S., & Leibowitz, A. (1996). The importance of child-care characteristics to

choice of care. *Journal of Marriage and Family, 58*, 759-773.

Karoly, L., Kilburn, M.R., & Cannon, J. (2005). *Early childhood interventions: Proven*

results, future promise. Santa Monica, CA: RAND Corporation.

Kirby, G., Boller, K., & Tout, K. (2010). Virginia Star Quality Initiative: QRS Profile.

Retrieved from: http://www.mathematica-

mpr.com/publications/*pdfs/earlychildhood /virginia.pdf*

Kontos, S., Howes, C., Shinn, M., & Galinsky, E. (1995). *Quality in family child care*

and relative care. New York: Teachers College Press.

LaParo, K.M., Hamre, B.K., & Pianta, R.C. (2011). *Classroom Assessment Scoring*

System- Toddler (Toddler CLASS). Baltimore: Brookes.

LaParo, K.M., Hamre, B.K., Locasale-Crouch, J., Pianta, R.C., Bryant, D., et al. (2009).

Quality in kindergarten classrooms: Observational evidence for the need to

increase children's learning opportunities in early education classrooms. *Early*

Education and Development, 20(4), 657-692.

Lazarsfeld, P.F. (1954). *Mathematical thinking in the social sciences.* New York, NY:

Free Press.

LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., Bryant, D.,
  Clifford, R., Early, D., & Barbarin, O. (2007). Observed classroom quality
  profiles in state-funded pre-kindergarten programs and associations with teacher,
  program, and classroom characteristics. *Early Childhood Research Quarterly, 22*,
  3-17.

Lugo-Gil, J., Sattar, S., Ross, C., Boller, K., Tout, K., & Kirby, G. (2011). *The quality
  rating improvement system evaluation toolkit.* OPRE Report #2011-31.
  Washington, DC: U.S. Department of Health and Human Services,
  Administration for Families, Office of Planning, Research, and Evaluation.

Mashburn, A.J. (2008). Quality of social and physical environments in preschools and
  children's development of academic, language, and literacy skills. *Applied
  Developmental Science, 12*(3), 113-127.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., et
  al. (2008). Measures of classroom quality in prekindergarten and children's
  development of academic, language and social skills. *Child Development, 79*,
  732–749.

McCarton, C.M., Brooks-Gunn, J., Wallace, I.F., Bauer, C.R., Bennett, F.C., Bernbaum,
  J.C., & Meinert, C.L. (1997). Results at age 8 yearsof early intervention for low-
  birth-weight premature infants. *Journal of the American Medical Association,
  277*, 126-132.

McLachlan, G.J., & Peel, D. (2000). *Finite mixture models.* New York, NY: Wiley.

Minnesota Department of Human Services. (2011). An early childhood quality rating and

improvement system: A report to the Minnesota legislature. Retrieved from:

http://archive.leg.state.mn.us/docs/2011/mandated/110352.pdf

Mitchell, A. (2005). *Stair steps to quality: A guide for states and communities developing*

quality rating systems for early care and education. Alexandria, VA: United Way

Success by 6.

Mitchell, A. (2009). *Quality rating improvement systems as a framework for early care*

and education system reform. Boston, MA: BUILD Initiative.

Muthén, B.O. (2002). Beyond SEM: General latent variable modeling. *Behaviormetrika,*

29, 81-117.

Muthén, B.O., & Muthén, L.K. (1998-2010). *Mplus version 6.0 user's guide.* Los

Angeles, CA: Author.

Muthén, B., & Muthén, L. (2000). Integrating person-centered and variable-centered

analyses: Growth mixture modeling with latent trajectory classes. *Alcoholism:*

Clinical and Experimental Research, 24, 882-891.

NICHD ECCRN. (1996). Characteristics of infant child care: Factors contributing to

positive caregiving. *Early Childhood Research Quarterly, 11*, 269-306.

NICHD ECCRN. (2002). Child-care structure→Process→Outcome: Direct and indirect

effects of child-care quality on young children's development. *Psychological*

Science, 13(3), 199–206.

Norris, D. J., Dunn, L., Eckert, L. (2003). *"Reaching for the stars": Center validation*

study final report. Norman, OK: Early Childhood Collaborative of Oklahoma.

Nylund, K., Bellmore, A., Nishina, A., & Graham, S. (2007). Subtypes, severity, and

structural stability of peer victimization: What does latent class analysis say?

Child Development, 78, 1706-1722.

Paulsell, D., Tout, K., & Maxwell, K. (2013). Evaluating implementation of quality

rating improvement systems. In T. Halle, A. Metz, & I. Martinez-Beck. (Eds.),

Applying implementation science in early childhood programs and systems (pp.

269-293). Baltimore, MD: Brookes.

Peisner-Feinberg, E., & Burchinal, M. (1997). Relations between preschool children's

child care experiences and concurrent development: The Cost, Quality and

Outcomes Study. *Merrill-Palmer Quarterly, 43*(3), 451-477.

Peisner-Feinberg, E.S., Burchinal, M., Clifford, R.M., Culkin, M., Howes, C., Kagan,

S.L., & Yazejian, N. (2001). The relation of preschool child care quality to

children's cognitive and social developmental trajectories through second grade.

Child Development, 72(5), 1534-1553.

Peisner-Feinberg, E.S., & Yazejian, N. (2010). Research on program quality. In P.W.

Wesley & V. Buysse (Eds.), *The quest for quality* (pp. 21-45). Baltimore, MD:

Brookes.

Phillips, D., & Howes, C. (1987). Indicators of quality in child care: Review of research.

In D. Phillips (Ed.), *Quality in child care: What does research tell us?* (pp. 1-19).

Washington, DC: National Association for the Education of Young Children.

Phillips, D., Mekos, D., Scarr, S., McCartney, K., & Abbott-Shim, M. (2000). Within and

beyond the classroom door: Assessing quality in child care centers. *Early

Childhood Research Quarterly, 15*(4), 475-496.

Pianta, R. C. (2001). Student– Teacher Relationship Scale: Professional manual. Lutz,

FL: Psychological Assessment Resources.

Pianta, R.C., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., et al. (2005).

Features of pre-kindergarten programs, classrooms, and teachers: Do they predict

observed classroom quality and child-teacher interactions? *Applied*

Developmental Science, 9(3), 144-159.

Pianta, R.C., LaParo, K.M., & Hamre, B.K. (2008). *Classroom Assessment Scoring*

System (CLASS). Baltimore: Brookes.

QRIS National Learning Network. (2013). Current status of QRIS in states. Retrieved

from: http://www.qrisnetwork.org/qris-state-contacts-map

Rock, D. A., & Pollack, J. M. (2002). Early Childhood Longitudinal Study—

Kindergarten Class of 1998–1999, Psychometric report for kindergarten through

first grade. NCES 2002-05.Washington, DC: National Center for Education

Statistics.

Sabol, T.W., Pianta, R.C., Downer, J., & Cao, Y. (2011, March). Improving Child Care

Quality in Virginia: An Evaluation of the Star Quality Initiative. Paper presented

at the Biennial Meeting of the Society for Research in Child Development,

Montreal, Quebec, Canada.

Schaack, D., Tarrant, K., Boller, K., & Tout, K. (2012). Quality rating improvement

systems: Frameworks for early care and education systems change. In S.L. Kagan

& K. Kauerz (Eds.), *Early Childhood Systems: Transforming Early Learning* (pp.

71-86). New York: Teachers College Press.

Schilder, D., Young, J., Anastasopoulous, L., Kimura, S., & Rivera, B. (2011). *Massachusetts quality rating and improvement system provisional standards study: Final report.* Boston, MA: Massachusetts Department of Early Education and Care.

Schwartz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6,* 461-464.

Schweinhart, L.J., Barnes, H.V., Weikart, D.P., Barnett, W.S., & Epstein, A.S. (1993). *Significant benefits: The High/Scope Perry Preschool study through age 27.* Ypsilanti, MI: High/Scope Press.

Sclove, S.L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52,* 333-343.

Seo, S. (2003). Early child care choices: A theoretical model and research implications. *Early Child Development & Care, 173,* 637-650.

Shonkoff, J.P., & Phillips, D.A. (2000). *From neurons to neighborhoods: The science of early childhood development.* Report of the National Research Council. Washington, DC: National Academies Press.

Smith, M..W., Brady, J.P., & Anastasopoloulos, L. (2008). *Early Language and Literacy Classroom Observation Tool Pre-K (ELLCO Pre-K).* Baltimore: Brookes.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). Preventing reading difficulties in young children.Washington, DC: National Academy Press.

Swenson-Klatt, D. & Tout, K. (2011). Measuring and rating quality: A state perspective on the demands for quality measurement in a policy context. In M. Zaslow, I.

Beck, K. Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings.* Baltimore, MD: Brookes.

Tout, K. & Maxwell, K. (2010). Quality rating and improvement systems: Achieving the promise for programs, families, and early childhood systems. In P.W. Wesley & V. Buysee (Eds.), *The quest for quality: Promising innovations for early childhood programs.* Baltimore, MD: Brookes.

Tout, K. & Starr, R. (2013). Key elements of a QRIS validation plan: Guidance and planning template. OPRE 2013-11. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G., & Boller, K. (2010). *Compendium of quality rating systems and evaluations.* Produced for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Tout, K., Zaslow, M., Halle, T., & Forry, N. (2009). *Issues for the next decade of quality rating and improvement systems* (Issue Brief). Administration for Children and Families, Office of Planning, Research and Evaluation and Child Trends.

U.S. Department of Education, & U.S. Department of Health and Human Services. (2011). Race to the Top Early Learning Challenge Executive Summary. Retrieved from: http://www2.ed.gov/programs/racetothetop-earlylearningchallenge/exec-summ.pdf

U.S. Department of Health and Human Services. (2013). Retrieved from: http://www.hhs.gov/recovery/programs/acf/childcare.html

Vandell, D. L., & Wolfe, B. (2000). *Child care quality: Does it matter and does it need to*

 be improved? Institute for Research on Child Poverty: University of Wisconsin-

 Madison. Retrieved from: http://www.irp.wisc.edu/publications/sr/pdfs/sr78.pdf

Vandell, D. L., Belsky, J., Burchinal, M., Steinberg, L., & Vandergrift, N. (2010). Do

 effects of early child care extend to age 15 years? Results from the NICHD study

 of early child care and youth development. *Child Development, 81*(3), 737-756.

Virginia Department of Social Services. (2009). *Virginia Star Quality Initiative for*

 continuous quality improvement pilot- phase III (VDSS RFP No. OECD-09-063).

 Richmond, VA: VDSS.

Virginia Department of Social Services. (2013). *Virginia Quality Rating and*

 Improvement System (VDSS RFP No. OECD-13-054). Richmond, VA: VDSS.

Virginia Star Quality Initiative. (2009). Standards for classroom-based programs.

 Retrieved from:

 http://www.smartbeginnings.org/index.asp?Type=B_BASIC&SEC={630988B9-

 7C1B-48A1-8D5E-9807F5F55C1A}

Virginia Star Quality Initiative. (2012). The Virginia Star Quality Initiative: Guidelines

 and procedures manual. Richmond, VA: VECF.

Wesley, P.W., & Buysse, V. (2010). Changing times and the quest for quality. In P.W.

 Wesley & V. Buysse (Eds.), *The quest for quality* (pp. 1-19). Baltimore: Brookes.

Westervelt, G., Sibley, A., & Schaack, D. (2008). Quality in programs for young

 children. In S. Feeney, A. Galper, & C. Seefeldt (Eds.) *Continuing issues in early*

 childhood education. (3rd ed., pp. 83-98). Upper Saddle River, NJ: Pearson.

Whitebrook, M., Howes, C., & Phillips, D. (1990). *Who cares? Child care teachers and the quality of care in America.* Final report of the National Child Care Staffing Study. Oakland, CA: Child Care Employee Project.

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. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Zellman, G. L., Brandon, R. N., Boller, K., & Kreader, J. L. (2011). *Effective evaluation of quality rating and improvement systems for early care and education and school-age care,* Research-to-Policy, Research-to-Practice Brief OPRE 2011-11a. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Zellman, G.L., & Perlman, M. (2008). *Child care quality rating and improvement systems in five pioneer states: Implementation issues and lessons learned.* Santa Monica, CA: RAND Corporation.

Zellman, G., Perlman, M., Le, V. N., & Setodji, C. M. (2008). *Assessing the validity of the Qualistar early learning quality rating and improvement system as a tool for improving child-care quality.* Santa Monica, CA: RAND Corporation.

Zinzeleta, E., & Little, N. K. (1997). How do parents really choose early childhood programs? *Young Children, 52*(7), 81-93.

CURRICULUM VITA

Kathryn Squibb
3328 Old Kirkwood Dr.
Virginia Beach, VA 23452
757-206-6012
katie.squibb@gmail.com

EDUCATION:

Old Dominion University, Norfolk, VA
Ph.D. Candidate, Early Childhood Education, Expected degree date: August, 2013
Recipient of Doctoral Fellowship, Research Assistantship, Teaching Assistantship
Recipient of Child Care Research Scholars grant through the HHS Administration for
Children and Families
Dissertation topic: *Quality profiles in early childhood: An example from Virginia's
Quality Rating Improvement System*

Old Dominion University, Norfolk, VA
M.S.Ed., Early Childhood Education, August, 2007
Recipient of Teaching Assistantship

College of William and Mary, Williamsburg, VA
B.A., Linguistics, May, 2003

PROFESSIONAL EXPERIENCE:

Center for Advanced Study of Teaching and Learning, Charlottesville, Va. August
2012-present
Curriculum Developer

Virginia Early Childhood Foundation, Richmond, Va. June 2012-present
Research and Evaluation Associate

Virginia Star Quality Initiative, Richmond, Va. February 2011-present
Chair, Star Quality Advisory Team
Master Star Quality Rater

Tidewater Community College, Virginia Beach, Va. May 2011-present
Instructor

Center for Advanced Study of Teaching and Learning, Charlottesville, Va. February
2011-August 2011

PREP Research Coordinator

Virginia Early Childhood Foundation, Richmond, Va. February 2011-August 2011
VSQI Pilot Project Specialist

Old Dominion University, Norfolk, Va. May 2011-September 2011
Adjunct Instructor

ODU Research Foundation, Norfolk, Va. September 2009-July 2011
Literacy Coach Mentor, Early Reading First

Smart Beginnings South Hampton Roads, Norfolk, Va. May 2007-February 2011
QRIS Program Director
Community Engagement Coordinator

Old Dominion University, Norfolk, Va. August 2006-September 2009
Research Assistant
Teaching Assistant

APPOINTMENTS:

Reviewer, Race to the Top Early Learning Challenge, U.S. Departments of Education, Health and Human Services
Chair, Star Quality Advisory Team, Virginia Star Quality Initiative
Member, QRIS Board Committee, Virginia Early Childhood Foundation
Member, Virginia Infant Toddler Specialist Network Leadership Council, Child Development Resources
Member, TCC Child Development Lab Advisory Team, Tidewater Community College
Moderator, "Collaborative Relationships: Bridging the Transition from Early Care to Educational Settings", Old Dominion University State of Education Summit