

USING THE EARLY CHILDHOOD LONGITUDINAL STUDY BIRTH COHORT DATA
TO EXPLORE MEASUREMENT INVARIANCE FOR THE EARLY CHILDHOOD
ENVIRONMENT RATING SCALE-REVISED AND THE ARNETT CAREGIVER
INTERACTION SCALE

By

CRYSTAL DAWN BISHOP

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2013

© 2013 Crystal Dawn Bishop

Dedicated in loving memory to Ruby Crowe (aka Granny), for inspiring me to seek
higher education

ACKNOWLEDGEMENTS

When I recall the path that has led me to the pursuit and completion of a doctoral degree, the most prominent word that comes to mind to describe my experience is “serendipitous.” Over the past 10 years, a series of unexpected and extremely fortunate events has led me down a path that has nurtured my passions, challenged me to think critically about all the things I think I know, and strengthened my dedication to making meaningful contributions to young children with disabilities and their families. Throughout my journey, I have had the good fortune to be surrounded by people whose support has shaped my experience and my success indescribably.

I wish to thank my family for their unwavering support of my academic pursuits and seemingly endless changes in career paths. Even when they questioned my sanity, or the possibility that I would ever finish school and get a real job, they have continued to offer encouragement. Their assurances that home was only a phone call away and their confidence in my ability to do anything I wanted led me to move forward, undaunted, into uncharted territory several times over. I am especially grateful for their constant reminders to stay grounded in the midst of hours of theorizing.

I would like to thank my mother for instilling in me a love of books and learning, which has persevered over time. She has also taught me the beauty of a balanced life, and her ability to create fun in any situation has reminded me that sometimes you just have to play dress-up for a little while before you can go on with the serious stuff. I thank my father for teaching me that the most meaningful achievements are the result of hard labor and sacrifice. His work ethic is unrivaled, and I am proud to say he was one of my earliest teachers. He will never know how much he has taught me over the years, or how invaluable his lessons have been to my success. I would also like to thank my

brother for reminding me that higher education is a luxury, and for having more faith in me than I could ever have in myself.

In addition to my biological family, I have obtained an “academic” family, whose support, knowledge, and perspectives have guided me and challenged me to become a better scholar. My decision to pursue a doctoral degree at all is due, in large part, to the guidance provided by Dr. Ann Kaiser. She saw in me a passion, a love of research, and potential to continue my academic career, and she was the first to encourage me (strongly, and against my will) to consider applying to doctoral programs.

I have been fortunate to have the opportunity to learn from and work with brilliant scholars. I would like to express my deepest gratitude to my committee chair, Dr. Patricia Snyder. Pat has an uncanny knack for taking a poorly articulated vision and crafting it into a rich and deeply meaningful learning experience. Even in the midst of hundreds of “emergencies,” the interests of her students are at the center of her work. Because of Pat’s dedication to excellence as a scholar and mentor, the past 5 years of my life have been filled with “important growth opportunities,” each of which has challenged me to become a more critical thinker, helped me gain experience and competence, and reminded me that no matter how much I learn, there will always be more I don’t know. In addition to supporting my academic endeavors, Pat has supported me through numerous and unexpected life transitions, and I am thankful to have had her friendship and support through multiple changes in both identity and priorities. It has been an honor to learn from her and to work along-side her, and I look forward to continuing our partnership together.

I also thank my other committee members, Drs. Jamie Algina, Walter Leite, and Jean Crockett, for their support through this process. I thank Jamie and Walter for their eternal patience as I muddled my way through a minor in research methodology. Both of them have spent countless hours huddled around a computer, helping me understand the nuances of writing syntax, making sense of findings, and interpreting statistical equations into language I can decipher and relate to. Jean Crockett has challenged me to see “the big picture” and to be more thoughtful about what I want to be when I grow up. Although she was not part of my official doctoral committee, I would like to thank Dr. Beth Rous for agreeing, days before I submitted this dissertation for review by my committee, to read this work and provide feedback.

I would also like to thank a number of other people who have become a part of my “academic family.” Dr. Robert Crow (aka “Dr. Bob”) has been an enthusiastic partner in many of my academic pursuits, and, because of him, I will be forever striving to achieve “verbal precision.” Although my time knowing her was far shorter than I would have liked, I am appreciative of Pearl Snyder’s (aka “Mama Pearl’s”) presence along this journey. She was an expert observational coder, one of the only people I have ever met who could keep Dr. Bob on task, and her zest for life was both refreshing and contagious. I also thank Tara McLaughlin, Salih Rakap, Cathy Pasia, Cinda Clark, and Team Snyder for enduring multiple dramatizations regarding the dissertation process. I would especially like to thank Tara McLaughlin for always keeping her office door open, despite multiple interruptions every single day; Salih Rakap for reminding me that longer is not better; and Tiffany McMonigle for her positive attitude and constant encouragement.

Finally, I would like to thank my husband, Jesse Bishop, and my daughter, Kate. Jesse and Kate were the most unexpected and fortunate additions to my life during this journey. I met Jesse in the throes of some pretty stressful times, and he has been completely undaunted by my many variations of stress and craze. With his partnership, nothing seems outside the realm of possibility (for instance, having a baby, defending qualifying exams, writing a dissertation proposal, and buying a house, all within the span of one year). Although I doubt he would agree I have achieved these skills, he has helped me to spend more time focusing on what is going well in the present and to expect the best possible outcome for the future. He has also given me a constant reminder of why I chose this path in the first place. The arrival of our daughter in the middle of this journey has forever changed my perspective and solidified my devotion to my work. She reminds me every day the value of this work, and she has become the light that keeps me focused and fuels my passion.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGEMENTS	4
LIST OF TABLES	12
LIST OF FIGURES.....	14
LIST OF ABBREVIATIONS.....	15
LIST OF DEFINITIONS	16
ABSTRACT	18
CHAPTER	
1 INTRODUCTION	20
Context for the Study	24
Policy Context for the Study	24
Recommendations for Developmentally Appropriate Practice in ECE	26
Recommended Practices for Infants and Young Children with Special Needs.....	27
Conceptualizations of Quality in ECE	28
Use of Instruments Designed to Measure Quality in ECE	31
Examining the Quality of ECE Nation-Wide	31
Characterization of the Quality of ECE for Children with Special Needs....	33
Quality Rating and Improvement Systems.....	34
Summary of the Context for the Study	35
Statement of the Problem	36
Purpose of the Present Study	40
Conceptual Frameworks	41
Early Childhood Longitudinal Study (Birth Cohort) Conceptual Framework	42
Measurement Models	43
ECERS-R measurement model	44
CIS measurement model	46
ECLS-B Data Set.....	47
Research Questions	49
Significance of the Study	49
Delimitations	50
Limitations.....	51
Chapter Summary.....	54
2 REVIEW OF THE LITERATURE	59
Search Procedures	59

Search Procedures for Topics 1 and 2	60
Search Procedures for Topics 3 and 4	62
Summary of Sources Reviewed	63
Defining Quality in ECE	64
Measuring Quality in ECE	68
Approaches for Quantifying Quality in ECE	69
Content and Scope of Quality Measures in ECE	72
Establishing Validity Evidence for Measures of Quality in ECE	73
Empirical Studies Examining Measurement Invariance in ECE Programs	77
Early Childhood Environment Rating Scale-Revised	79
Development and Revision of the ECERS	80
Structural Validity Evidence for the ECERS/ECERS-R	83
Empirical studies examining the factor structure of the ECERS.....	84
Empirical studies examining the factor structure of the ECERS-R.....	86
Summary of empirical studies examining factor structure of the ECERS/ECERS-R	90
Use of ECERS/ECERS-R to Examine Quality in Classrooms Enrolling Children with Special Needs	92
Use of ECERS to examine the quality in ECE for children with special needs.....	94
Use of ECERS-R to examine quality in ECE classrooms for children with special needs	97
Summary of research examining the quality in ECE for children with special needs.....	101
Arnett Caregiver Interaction Scale	104
Development of the CIS	104
Empirical Studies Examining Factor Structure of the CIS	105
Use of the CIS to Characterize Process Quality in Inclusive Preschool Classrooms	108
Summary	110
3 METHODOLOGY	122
Research Questions	122
Research Design	122
ECLS-B Study and Data Set.....	124
ECLS-B Sampling Procedures	124
Sampling Procedures for Preschool Child Care Observations	125
ECLS-B Response Rates and Sampling Weights	127
Response rates for CCO subsample	128
ECLS-B Instrumentation.....	130
Birth Certificate Data	132
Parent Interviews.....	132
ECE Provider Interviews.....	133
Child Care Observations	134
ECERS-R.....	134
ECERS-R certification.....	136

ECERS-R data provided in the ECLS-B data file	138
Caregiver Interaction Scale.....	139
Caregiver Interaction Scale certification.....	140
Caregiver Interaction Scale data provided in the ECLS-B data file	141
Definition of Key Variables.....	142
Children with Special Needs.....	143
ECLS-B focal children with special needs.....	143
Non-focal children with special needs.....	144
INC and NSN classrooms	145
Analytic Sample for the Present Study	146
Analytic Procedures	147
Data File Preparation	148
Weighting	150
Analyses.....	151
Multiple group confirmatory factor analyses.....	151
Configural invariance model.....	152
Strict invariance	152
Group Comparisons of Latent Variables	153
Summary	154
4 RESULTS	159
Context for Reporting and Interpreting Findings	159
Research Questions 1 and 2	160
Estimation of Configural Invariance Models	161
Evaluating Model Fit for Configural Invariance Models	162
ECERS-R Model Fit	163
CIS Model Fit.....	163
Research Question 3	164
Estimation of Strict Invariance Models	164
Evaluating Model Fit for the Strict Invariance Models.....	165
ECERS-R Model Fit	166
CIS Model Fit.....	168
Research Question 4	169
Examining Group Differences in Latent Variable Scores.....	170
Mean differences in ECERS-R latent variables.....	171
Mean differences in CIS latent variables.....	171
Summary	172
5 DISCUSSION	186
Interpretation of Findings	187
Measurement Invariance of the ECERS-R.....	188
Factor Structure of the ECERS-R.....	190
Group Differences in ECERS-R Latent Variables.....	193
Activities and materials	194
Language and interactions.....	196

Personal care and safety	197
Measurement Invariance of the CIS	198
Factor Structure of the CIS.....	200
Group Differences in CIS Latent Variables.....	201
Implications from the Present Study	201
Practical Implications of Findings	202
Policy and Research Implications of Findings	205
Recommendations for Future Research	209
Summary	212

APPENDIX

A INSTRUMENT ITEMS AND MEASUREMENT MODELS	215
B INTERVIEW QUESTIONS USED TO IDENTIFY CHILDREN WITH SPECIAL NEEDS	224
C VARIABLE CODING SYNTAX.....	229
D ANALYTICAL SYNTAX.....	234
REFERENCE LIST.....	241
BIOGRAPHICAL SKETCH.....	254

LIST OF TABLES

<u>Table</u>		<u>page</u>
1-1	Percentage of Children Under Age 5 Attending Center-Based ECE Programs in 2011 by Employment Status and Characteristics of Mother	56
2-1	Summary of Content for Selected Observation-Based Instruments Used to Measure Dimensions of Quality in ECE.....	111
2-2	Summary of Empirical Studies Examining Factor Structure of the ECERS	113
2-3	Summary of Empirical Studies Examining the Factor Structure of the ECERS-R	114
2-4	Comparison of Factors Identified for the ECERS-R.....	115
2-5	Comparison Groups and Interpretations from Studies Examining the Quality in ECE in Classrooms where Children with Special Needs were Enrolled	119
3-1	Internal Consistency Reliability Scores for Published ECERS-R and CIS Subscales.....	155
3-2	Weighted Percentage of ECLS-B Focal Children in INC and NSN classrooms	156
3-3	Percentage of Center-Based Care Types by Classroom Group Classification.	157
3-4	Mean Scores on Published ECERS-R and CIS Subscales	158
4-1	Model Fit Statistics for Multiple Group Confirmatory Factor Models	174
4-2	Non-Standardized Factor Loadings from the ECERS-R Strict Invariance Model.....	175
4-3	Standardized Factor Loadings for INC Classrooms from the ECERS-R Strict Invariance Model	176
4-4	Internal Consistency Reliabilities for ECERS-R Latent Variable Subscales	177
4-5	Percentage of Scores for Each ECERS-R Score Category in NSN Classrooms.....	178
4-6	Percentage of Scores for Each ECERS-R Score Category in INC Classrooms.....	179
4-7	Non-Standardized Factor Loadings from the CIS Strict Invariance Model.....	180
4-8	Standardized Factor Loadings INC Classrooms from the CIS Strict Invariance Model	181

4-9	Internal Consistency Reliabilities for CIS Latent Variable Subscales	182
4-10	Percentage of Scores for Each CIS Score Category in NSN Classrooms.....	183
4-11	Percentage of Scores for Each CIS Response Category in INC Classrooms. .	184
4-12	Mean Scores for ECERS-R and CIS Latent Variable Subscales.....	185
A-1	Subscales, Items, and Proposed Factors for the ECERS-R.....	216
A-2	Arnett Scale of Caregiver Behavior Items and Item Wording in the ECLS-B Child Care Observations	220

LIST OF FIGURES

<u>Figure</u>		<u>page</u>
1-1	ECLS-B Conceptual Framework.....	57
1-2	Conceptual Model of Relationships Between Global, Structural, and Process Quality and Child Outcomes	58
A-1	Measurement Model for the ECERS-R.....	218
A-2	Bi-factor Measurement Model for the CIS.	223

LIST OF ABBREVIATIONS

CIS	Arnett Caregiver Interaction Scale
CFA	Confirmatory factor analysis
NSN	No special needs
ECE	Early care and education
ECLS-B	Early Childhood Longitudinal Study—Birth Cohort
ECERS	Early Childhood Environment Rating Scale
ECERS-R	Early Childhood Environment Rating Scale-Revised
IDEA	Individuals with Disabilities Education Improvement Act
INC	Inclusive
MG-CFA	Multiple group confirmatory factor analysis
QRIS	Quality rating and improvement system

LIST OF DEFINITIONS

Center-based	Early care and education program operated in a location that is not someone's home.
Children with special needs	Children whose data in the ECLB-S B data set met one or more of the following criteria: (a) had received early intervention or preschool special education services; (b) had been diagnosed with conditions that would make them eligible for early intervention or preschool special education services or that were likely to be associated with developmental delays (e.g., sensory impairment, chromosomal abnormality); (c) had suspected delays (e.g., difficulty using arms or legs), as reported by parents or professionals, or (d) were born with a birth weight below 1500 g.
Classrooms with no special needs (NSN)	Refers to classrooms in which children in the ECLS-B sample were enrolled that, at the time of ECLS-B data collection, only enrolled children without special needs .
Early care and education classroom	A classroom located within a home- or center-based early care and education program where a child is cared for or taught.
Distal quality indicators	Indicators of early care and education quality not experienced directly by the child, but that have an indirect influence on children's health, development, and learning (Dunn, 1993).
Early care and education program	A home- or center-based program where children receive routine care and education by persons other than their parents, primary guardians, or relatives.
Global quality	Overall quality of an early care and education program or classroom (Phillips & Howes, 1987).
Home-based	An early care and education program operated in someone's home.
Inclusive classrooms (INC)	Refers to classrooms in which children in the ECLS-B sample were enrolled that, at the time of ECLS-B data collection, met the following criteria: (a) both children with and without special needs were enrolled, and (b) 75% or less of the children enrolled in the classroom had special needs.
Inclusive environment	An early learning environment (e.g., ECE classroom, ECE program) in which both children with and without special needs are enrolled.

Measurement invariance	The extent to which an instrument yields measurements of the same attribute when applied under different conditions (Horn & McArdle, 1992).
Preschool program	An early care and education program in which children ages 3 to 5 are enrolled.
Process quality	Dynamic processes or interactions between individuals within an early care and education program or classroom (Cassidy, Hestenes, Hansen, et al., 2005).
Proximal quality indicators	Indicators of early care and education quality that are experienced directly by the child (Dunn, 1993).
Structural quality	Inputs to dynamic processes or interactions between individuals that are independent of human interaction (Cassidy, Hestenes, Hansen et al., 2005).

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

**USING THE EARLY CHILDHOOD LONGITUDINAL STUDY BIRTH COHORT DATA
TO EXPLORE MEASUREMENT INVARIANCE FOR THE EARLY CHILDHOOD
ENVIRONMENT RATING SCALE-REVISED AND THE ARNETT CAREGIVER
INTERACTION SCALE**

By

Crystal Dawn Bishop

December 2013

Chair: Patricia Snyder
Major: Special Education

Quality in early care and education (ECE) has become a national priority.

Growing interest in characterizing quality in ECE has resulted in a necessity for valid measures of quality that are applicable across different types of ECE settings. The Early Childhood Environment Rating Scale-Revised (ECERS-R) and the Arnett Caregiver Interaction Scale (CIS) are two instruments used to characterize different dimensions of quality in ECE for numerous purposes, including examining quality in ECE classrooms and programs where children with special needs are enrolled. Although there have been some studies providing structural validity evidence for these instruments, there have been no empirical studies conducted to examine measurement invariance for either instrument across center-based ECE classrooms with varying compositions of children with special needs.

The present study involved secondary analyses of data from a large sample of center-based ECE classrooms observed as part of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B). The purpose of this study was to examine measurement

invariance for the ECERS-R and the CIS across two types of classrooms: inclusive (INC) classrooms, in which both children with and without special needs were enrolled, and classrooms in which no children with special needs were enrolled (NSN). To examine measurement invariance, multiple group confirmatory factor analyses were conducted separately for each instrument. For each instrument, there was strong evidence of measurement invariance across INC and NSN classrooms.

Regression analyses were conducted to examine group differences in the quality of ECE across INC and NSN classrooms. Findings suggested higher quality activities and materials and personal and safety practices in INC classrooms. Although there was evidence to suggest each instrument measures the quality of language and interactions occurring in INC and NSN classrooms, neither instrument appeared to measure this dimension of quality with enough specificity or sensitivity to detect variation across the two types of classrooms studied.

Relevant background information, including a review of literature, is presented to situate the need for the study. A description of the methodologies used to conduct the study is provided, followed by the presentation and interpretation of findings. Implications for practice, policy, and research are discussed.

CHAPTER 1 INTRODUCTION

Over the past 30 years, the quality of early care and education (ECE) has become a topic of major concern for researchers, policymakers, and early childhood practitioners. During this time, the number of children under 5 years of age who attend some type of ECE program (e.g., home-based programs, Head Start programs, public or private preschool programs, child care programs, nursery schools) has risen steadily (Laughlin, 2013). Concomitant with the increases in the numbers of children enrolled in ECE programs has been a growing number of research studies designed to characterize the quality of ECE programs in the United States (e. g., Helburn, 1995; Layzer, Goodson, & Moss, 1993; Ruopp et al., 1975, Whitebook, Phillips, & Howes, 1989;) and to examine relationships between various dimensions of quality in ECE and young children's health and early development (e.g., Bryant, Burchinal, Lau, & Sparling, 1994; Burchinal et al., 2008; National Institute of Child Health and Human Development Early Child Care Research Network [NICHD ECCRN] 1998, 2002, 2003; Peisner-Feinberg & Burchinal, 1997; Peisner-Feinberg et al., 2001; Vogel, Xue, Moiduddien, Kisker, & Carlson, 2010). Findings from these studies have suggested the quality of ECE programs in the United States is, on average, only minimally adequate to promote children's health and development, but that high-quality ECE experiences are positively associated with children's cognitive, academic, and social-emotional development.

A number of instruments have been designed to measure quality in ECE programs or classrooms and to quantify key dimensions of quality in ECE (Halle, Whittaker, & Anderson, 2010). The key dimensions of quality in ECE typically operationalized on existing instruments are process and structural quality. Broadly

defined, process quality refers to the dynamic processes and interactions occurring in ECE programs and classrooms (Cryer, 1999; Phillips & Howes, 1987; Vandell & Wolfe, 2000). Structural quality refers to the “inputs” to process quality, which provide a framework in which dynamic processes operate (Cryer, 1999; Phillips & Howes, 1987). Global quality refers to the overall quality of an ECE program or classroom and is often quantified by combining measures of process and structural quality (Cryer, 1999; Phillips & Howes, 1987, Vandell & Wolfe, 2000). Although these definitions are often applied to describe program-wide quality, many instruments designed to measure quality in ECE are administered at the classroom level rather than at the program level (e.g., Arnett, 1989; Harms, Clifford, & Cryer, 1998, 2005; Pianta, La Paro, & Hamre, 2008, Smith, Dickinson, Sangeorge, & Anastasopoulos, 2002). One reason for this is that many of indicators of quality in ECE that are of particular interest to researchers, policymakers, and practitioners are experienced directly by children and are observable within the context of ongoing classroom activities and routines (e.g., adult-child interactions, child-child interactions, provision of activities and materials, personal care routines; Snow & Van Hemel, 2008).

A number of instruments that involve direct observations have been developed to quantify dimensions of quality in ECE (Halle et al., 2010; Snow & Van Hemel, 2008). Many of these instruments have been used for multiple purposes, such as: (a) self-assessment tools to inform quality improvement efforts, (b) dependent measures in research studies, (c) descriptive measures in research studies, and (d) assessments for external evaluations (Snow & Van Hemel, 2008). Many of the existing instruments designed to measure quality in ECE, however, were not initially developed for nor have

they been validated for each of these purposes. In addition, given the diversity of ECE programs and classrooms in the United States, the availability of instruments that have been validated for use across different types of ECE programs and classrooms is important.

Of particular relevance to the present study is the need to establish measurement invariance for instruments designed to quantify dimensions of classroom quality (Horn & McArdle, 1992; Horn, McArdle, & Mason, 1983; Vandenberg & Lance, 2000). Measurement invariance refers to the extent to which instruments yield comparable measurements of a particular attribute when administered under different conditions (Horn & McArdle, 1992). Measurement invariance has been described as essential to making meaningful interpretations of scores from instruments that have been administered under varying conditions, but this type of validation is often neglected (Horn & McArdle, 1992; Vandenberg & Lance, 2000).

The purpose of the present study was to explore measurement invariance as a form of validity evidence for two instruments that often have been used in research to characterize key dimensions of quality across a variety of early care and education (ECE) programs and classrooms. Data for the present study were obtained from the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B). Secondary analyses using the ECLS-B data were conducted to explore measurement invariance of the Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms et al., 1998) and the Arnett Caregiver Interaction Scale (CIS; Arnett, 1989) across two types of classrooms: (a) center-based preschool classrooms in which children with special needs were enrolled along with children without special needs (INC) and (b) center-based preschool

classrooms in which no children with special needs (NSN) were enrolled. For the present study, INC classrooms were defined as center-based classrooms that met the following criteria: (a) both children with and without special needs were enrolled at the time of ECLS-B data collection, and (b) 75% or less of the enrollment were children with special needs. NSN classrooms were defined as center-based classrooms in which there were no children with special needs enrolled at the time of ECLS-B data collection.

INC and NSN classrooms were chosen as the unit of analysis for the present study because they represent two types of ECE classrooms for which the quality of ECE is of interest to researchers, policymakers, and practitioners. A growing number of research studies have been conducted in which either the ECERS-R or the CIS has been used to compare key dimensions of quality across classrooms or programs with differing configurations of children with and without special needs (e.g., Buysse, Wesley, Bryant, & Gardner, 1999; Grisham-Brown, Cox, Gravil, & Missall, 2010; Hestenes, Cassidy, Shim, & Hegde, 2008; Knoche, Peterson, Edwards, & Jeon, 2006).

In this chapter, the context for the present study and a statement of the problem are provided to situate the need for the study. The purpose, conceptual frameworks, and information about ECLS-B are described. The research questions that guided the secondary analyses conducted for the present study are presented. A description of the significance of the study is provided to highlight the relevance of the findings for early childhood researchers, policymakers, and practitioners. Delimitations and limitations of the present study are presented followed by a brief summary.

Context for the Study

A number of contextual factors are relevant to the present study. Changes in early childhood policy and recommendations for practice in ECE are described to provide contextual information about how the implementation and utilization of ECE programs has evolved over the past 30 years. Early and contemporary conceptualizations of quality in ECE are presented to provide a context for the ways in which the quality of ECE in ECE classrooms and programs has been studied and characterized. Finally, an overview of the ways in which measures of quality in ECE have been used to characterize quality in ECE programs and classrooms is presented.

Policy Context for the Study

In the most recent United States census report, 61% of children under 5 years of age were reported to receive regular child care (Laughlin, 2013). Regular child care was defined as an arrangement that was used at least once per week (U.S. Dept. of Commerce, United States Census Bureau, 2013). Approximately 38% of these children attended center-based ECE programs (e.g., child care centers, Head Start programs, public or private preschools). As shown in Table 1-1, children who attended center-based ECE programs are diverse with respect to family structure, race, parental employment, and income level. In addition to differences in their demographics, children who attend center-based ECE programs have diverse abilities. The enrollment of young children with diverse abilities in center-based ECE programs is due, in large part, to federal mandates requiring ECE programs to make accommodations for children with or at risk for disabilities or developmental delays. Across these pieces of legislation, the terms “disability,” “at-risk,” and “developmental delay” are defined differently. For the present study, the term “children with special needs” is used to

encompass the variety of ways in which children with or at risk for developmental delays or disabilities have been described.

The Individuals with Disabilities Education Improvement Act (IDEA; 2004) specifies that children 3 to 21 years of age with disabilities who are identified as eligible to receive special education and related services must be provided a free appropriate public education (FAPE) in the least restrictive environment (LRE), which includes home- and center-based ECE programs. Recent data pertaining to children ages 3 through 5 who were eligible to receive special education and related services under the IDEA show approximately 51% of these children attended a home- or center-based ECE program rather than receiving services in a special education program, their home, or a separate service provider or location for more than 10 hours per week (Data Accountability Center, 2011).

In addition to the IDEA (2004), there are a number of other pieces of federal legislation contributing to the enrollment of young children with diverse abilities in center-based ECE programs. The Head Start Act (2007) mandates the provision of ECE for young children whose families meet specified income requirements, and requires 10% of children served by Head Start and Early Head Start programs to be children with disabilities, regardless of whether the income requirements are met. The Child Care and Development Block Grant of 1990 provides subsidies to increase access to high-quality ECE programs for families who meet low-income criteria. The use of these subsidies is permitted for families of children with disabilities up to age 19. Furthermore, federally funded ECE programs are prohibited from denying services to

children with disabilities, unless such services cause undue financial burden to the program (Americans with Disabilities Act, 2008).

Recommendations for Developmentally Appropriate Practice in ECE

The need for ECE programs to accommodate children with special needs has been reflected in changes in recommendations for developmentally appropriate practice in ECE. The National Association for the Education of Young Children's (NAEYC) position statements on developmentally appropriate practice and the accompanying guidelines for developmentally appropriate practice in ECE programs serving children from birth through age 8 (Bredekamp, 1987; Bredekamp and Copple, 1997; Bredekamp & Copple, 2009) have guided the design and evaluation of ECE programs nationwide since their adoption in 1986. The original guidelines (Bredekamp, 1987) were adopted at the same time as federal legislation requiring FAPE in the LRE for children with or at risk for disabilities was amended to include young children from birth through age 5 (P.L. 99-457) and emphasized the importance of child-centered activities in which children directed their own learning through play, with limited instruction from adults. Although the NAEYC guidelines were originally targeted toward ECE programs serving children without special needs, many advocated for their adoption with respect to ECE programs serving children with special needs (e.g., Berkeley & Ludlow, 1989). The original guidelines for developmentally appropriate practice were revised to reflect changes in ECE programs, which occurred in the decade following the passage of P. L. 99-457. The revised guidelines (Bredekamp & Copple, 1997) referred specifically to the "trend toward full inclusion of children with disabilities" (p. 4) in ECE programs. Revisions to the original guidelines included a greater emphasis on adapting instruction to meet the needs of individual children and providing a balance of child-directed and

adult-directed teaching (Bredekamp & Copple, 1997). The most recent revision of the guidelines for developmentally appropriate practice (Copple & Bredekamp, 2009) highlights even further the importance of individualizing learning experiences for children with diverse abilities and special needs. Specifically, these guidelines reflect the current political context for accountability in ECE, the implications of this context for ECE programs, and the importance of high-quality ECE programs in closing achievement gaps for children with special needs who are at biologic or environmental risk for later developmental delays. Special attention is focused on providing intentional and differentiated instruction to children based on their individual needs.

Recommended Practices for Infants and Young Children with Special Needs

In the years following the passage of P.L. 99-457 and the adoption of the original guidelines for developmentally appropriate practice, efforts were directed toward identifying recommended practices for infants and young children with special needs and their families (Division for Early Childhood [DEC] Task force on Recommended Practices, 1993; Odom & McLean, 1996; Sandall, McLean, & Smith, 2000; Sandall, Hemmeter, Smith, & McLean, 2005). Despite growing empirical support for the benefits of including of children with special needs in ECE programs (Buysse & Bailey, 1993), researchers in early intervention and early childhood special education stressed that the benefits incurred by the inclusion of young children with special needs in ECE programs alongside children without special needs were conditional on the quality of the ECE program and the manner in which ECE programs were implemented and operated (Lamorey & Bricker, 1993; Odom & McEvoy, 1988; Wolery & Wilbers, 1994). The most recent DEC recommended practices (Sandall, Hemmeter, Smith, & McLean, 2005) emphasize the provision of individualized ECE and related services for young children

with special needs in inclusive ECE programs. In 2009, NAEYC and DEC issued a joint position statement on early childhood inclusion, which emphasized the importance of providing high-quality ECE for all children in inclusive early learning environments and providing recommendations for improving ECE services. The position statement highlighted three key components of high-quality, inclusive ECE: (a) access for all children to a broad range of learning opportunities, activities, and environments; (b) full participation of all children with peers and adults in early learning activities; (c) and system-wide infrastructure supports to provide inclusive ECE services (NAEYC/DEC, 2009).

Conceptualizations of Quality in ECE

The brief background review suggests that, over the past 30 years, early childhood legislation along with developmentally appropriate and recommended practices have resulted in a burgeoning interest in characterizing and improving quality in the variety of ECE programs and classrooms in which young children are enrolled. Quality in ECE has often conceptualized along a continuum from low to high quality and characterized by multiple quality indicators that are either proximal or distal in nature (Dunn, 1993; Love, Schochet, & Meckstroth, 1996).

Proximal quality indicators purportedly have a direct influence on the child's health, safety, development, and learning and can be operationally defined so they are often observable within the context of ongoing classroom activities and routines (Dunn, 1993; Snow & Van Hemel, 2008). Examples of proximal quality indicators include the adequacy of the physical space to promote the health and safety of children, the provision of materials and activities sufficient to promote children's healthy development, and the interactions between adults and children in the ECE program.

(Cryer, 1999; Dunn, 1993; Love et al., 1996; Phillips & Howes, 1987; Vandell & Wolfe, 2000). Proximal quality indicators have been emphasized in guidelines for developmentally appropriate practice in ECE programs serving children ages birth through 8 (Bredekamp, 1987; Bredekamp & Copple, 1997; Copple & Bredekamp, 2009) and the DEC recommended practices for young children with or at risk for disabilities (Sandall, et al., 2005).

Distal quality indicators are often described as aspects of quality not experienced directly by the child or that, even if they are experienced directly, have an indirect influence on children's health, safety, development, and learning. Distal quality indicators historically have included information about the education and experiences of ECE administration and staff, staff wages, adult-child ratios, and group size (Cryer, 1999; Dunn, 1993; Phillips & Howes, 1987; Vandell & Wolfe, 2000). Recent developments in implementation science suggest additional distal quality indicators might include professional development, staff selection, facilitative administration, data-based decision making, and program leadership (Metz & Bartley, 2012; Metz, Halle, Bartley, & Blasberg, 2013). A number of studies have documented the importance of distal quality indicators as having an indirect influence on children's child's health, safety, development, and learning (e.g., Burchinal, Cryer, Clifford, & Howes, 2002; Clarke-Stewart, Vandel, Burchinal, O'Brien, & McCartney, 2002; Early et al., 2006; Howes, Phillips, & Whitebook, 1989).

The definitions for key dimensions of quality in ECE and their relationships to proximal and distal quality indicators have evolved over time. The key dimensions of quality in ECE typically described and measured by existing instruments are process

and structural quality. Process quality is broadly defined as the dynamic processes or interactions occurring in the ECE program or classroom, whereas structural quality is conceptualized as inputs to process quality (Cryer, 1999; Love et al., 1996; Phillips & Howes, 1987; Vandell & Wolfe, 2000). Global quality refers to the overall quality of the ECE program or classroom and is conceptualized as the combination of quality indicators related to structural or process quality in addition to indicators related to health and safety practices (Cryer, 1999; Phillips & Howes, 1987; Vandell & Wolfe, 2000). Proximal quality indicators typically have been conceptualized as indicators of process quality, and distal quality indicators typically have been conceptualized as indicators of structural quality. Although these conceptualizations of quality generally have been accepted and used by researchers and policymakers, there has been some disagreement regarding the categorization of all proximal quality indicators as indicative only of process quality (Cassidy, Hestenes, Hansen, et al., 2005; La Paro et al., 2012). It is generally agreed that proximal quality indicators related to interactions between adults and children are associated with process quality. Some researchers have suggested proximal quality indicators related to the arrangement of physical space, the provision of developmentally appropriate materials, and the provision of developmentally appropriate activities to support children's early development and learning also are indicators of process quality, because they are experienced directly by the child (e.g., Cryer, 1999; Phillips & Howes, 1987; Vandell & Wolfe, 2000). More recent conceptualizations of quality in ECE, however, have characterized these features of the classroom environment as indicators of structural quality, because they can be considered antecedents for more dynamic processes or interactions but their presence

alone does not guarantee the occurrence of dynamic processes or interactions (Cassidy, Hestenes, Hansen, et al., 2005).

Use of Instruments Designed to Measure Quality in ECE

Despite a lack of consensus over specific definitions for dimensions of quality in ECE, there has been a burgeoning interest in characterizing the quality of ECE in a variety of ECE programs and classrooms. Numerous instruments have been developed and used to characterize dimensions of quality in ECE classrooms and programs for a variety of purposes, including as dependent or descriptive measures of quality in research studies and as tools for self-assessment or external evaluations (Halle et al., 2010; Snow & Van Hemel, 2008). A brief review of the ways in which such instruments have been used to characterize quality in ECE is provided.

Examining the Quality of ECE Nation-Wide

Several large-scale research studies have been designed to characterize dimensions of quality in ECE programs and classrooms in the United States. The first study of this type, the National Day Care Study (Ruopp, Travers, Glantz, & Coelen, 1979), was focused primarily on characterizing day care quality by examining structural quality in 64 center-based ECE programs. The National Child Care Staffing Study (NCSS; Whitebook et al., 1989) was the first large-scale study in which observations of classrooms in children's ECE programs were conducted to permit characterization of process quality in addition to structural quality. The primary purpose for collecting these data was to examine relationships between structural and process quality. The Cost, Quality, and Child Outcomes Study (CQO; Helburn, 1995) and the Observational Study of Early Childhood Programs (OSEP; Layzer, Goodson, & Moss, 1993) also examined structural and process quality in center-based ECE programs. In addition to providing

descriptive information about the quality of center-based ECE programs, these studies examined relationships between different dimensions of quality in ECE and children's developmental outcomes. Subsequent large-scale studies in which there has been a focus on quality in ECE have generally examined quality descriptively in addition to examining relationships between dimensions of quality and children's short- and long-term developmental outcomes. Some of the large-scale studies conducted have been the Early Head Start Research and Evaluation Project (EHSRE; Office of Planning, Research, and Evaluation [OPRE], 1996-2010), the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B; Institute of Education Science [IES], National Center for Education Statistics [NCES], n. d.), the Head Start Family and Child Experiences Survey (FACES; OPRE, 1997-2013), and the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development (NICHD; National Institute of Health [NIH], 2012).

Findings from early large-scale studies suggested the quality of ECE programs or classrooms in the United States was mediocre at best (Helburn, 1995; Whitebook et al., 1989). Findings from studies in which the quality of ECE was related to children's developmental outcomes suggested ECE experiences characterized as high quality were related positively to children's short- and long-term developmental outcomes (e.g., Burchinal et al., 2008; NICHD ECCRN 1998, 2002, 2003; Peisner-Feinber & Burchinal, 1997; Peisner-Feinberg et al., 2001). Relationships between process quality and children's developmental outcomes have been found to be particularly important for children who are at biological (e.g., low birth weight, prematurity) or environmental (e.g., low income) risk for developmental delays. Such findings have provided empirical

evidence to support transactional and ecological conceptual models of early childhood development, which highlight the importance of interactions between environmental and biological factors to either support or hinder children's early development (e.g., Bronfenbrenner & Crouter, 1983; Bronfenbrenner & Morris, 1998; Horowitz, 1987, Sameroff & Chandler, 1975). Evidence available to date suggests experiences in ECE environments characterized as high quality can facilitate children's early learning and development and might buffer the adverse effects of biological (i.e., established medical conditions or disabilities, low birth weight, prematurity) or environmental (e.g., low income) vulnerabilities that impede learning and development (Horowitz, 1987; Rutter, 1987; Rutter et al., 2997; Shonkoff & Phillips, 2000). In contrast, exposure to low-quality ECE environments can create additional risk or vulnerability for children with biological or other environmental risk factors or vulnerabilities.

Characterization of the Quality of ECE for Children with Special Needs

A growing number of studies have focused on characterizing global, structural, and proximal quality in center-based ECE preschool programs or classrooms where young children with special needs are enrolled. The purpose of these studies has been to compare the quality of ECE across classrooms or programs with different compositions of children with and without special needs. Measures of quality dimensions for these studies were based on observations conducted in ECE classrooms. One of the earliest studies of quality in ECE programs or classrooms where children with special needs were enrolled was focused on comparing quality across ECE programs designed primarily for children with special needs and ECE programs designed primarily for children without special needs (Bailey, Clifford, & Harms, 1982). Other studies have focused on comparing dimensions of quality across

ECE classrooms in which only children with special needs were enrolled and classrooms in which both children with and without special needs were enrolled (e.g., La Paro, Sexton, & Snyder, 1998) or ECE classrooms or programs in which both children with and without special needs were enrolled and ECE classrooms or programs in which no children with special needs were enrolled (e.g., Buysse et al., 1999; Grisham-Brown et al., 2010; Hestenes et al., 2008; Knoche et al., 2006). More recent studies have used measures of classroom quality to make inferences about the quality of ECE received by a particular child or children enrolled in the classroom (Clawson and Luze, 2008; Wall et al., 2006). The shift in foci in the studies across time is reflective of changes in early childhood policy and recommendations for practice in ECE, which support the provision of ECE services in inclusive environments. In addition, these studies illustrate how instruments designed to measure the quality of ECE provided at the classroom level have been used to make inferences about differences in quality across different types of ECE programs and classrooms as well as for making inferences about the quality of ECE experienced by individual children enrolled in the classroom.

Quality Rating and Improvement Systems

In the past 15 years, there has been a heightened interest in the design and implementation of statewide systems for continuous monitoring and improvement of quality in ECE. A number of states have adopted Quality Rating and Improvement Systems (QRIS), which are designed to assess the quality of ECE provided by ECE programs and to communicate quality ratings to consumers (Tout, Starr, Soli, & Moddie, 2010; Tout, Zaslow, Halle, & Forry, 2009). Common features of QRIS include (a) standards of quality, which provide the basis for program quality ratings; (b) a process

for monitoring quality standards, which often includes the administration of an observational instrument designed to characterize quality in ECE classrooms or programs; (b) a process for supporting programs in quality improvement endeavors; (d) financial incentives to promote participation in QRIS; and (e) dissemination of program ratings to consumers (Mitchell, 2005; Zellman & Perman, 2008; Tout et al., 2009; Tout et al; 2010). The majority of QRIS involve the use of observational tools to assess the quality of ECE provided by programs (Tout et al., 2010). Typically, the quality rating process involves sampling ECE classrooms within a program to be observed, and aggregating their scores along with other sources of evidence (e.g., licensing compliance, group size and ratio requirements) to determine an overall quality rating for the program. In addition, some QRIS include a self-assessment component, which either requires or encourages programs to utilize instruments designed to asses quality in ECE conduct a self-evaluation of quality within their programs. Despite similarities in the instruments selected to measure quality in ECE for QRIS, a number of variations in their use have been noted. For example, QRIS differ in the number of classrooms sampled to acquire a program rating. In addition, there are variations in the procedures use to summarize scores from observations and to convert these scores to program quality ratings.

Summary of the Context for the Study

Numerous changes in the policies governing the implementation of ECE, guidelines for developmentally appropriate practice in ECE programs serving children from birth through age 8 (Bredekamp, 1987; Bredekamp and Copple, 1997; Bredekamp & Copple, 2009), and recommendations for practice for infants and young children with special needs and their families (DEC Task force on Recommended Practices, 1993;

Odom & McLean, 1996; Sandall, McLean, & Smith, 2000; Sandall, Hemmeter, Smith, & McLean, 2005) have resulted in a heightened interest in characterizing quality in ECE. Conceptualizations of ECE quality have evolved over time, and a number of instruments have been developed to measure dimensions of quality in ECE. Often, inferences regarding quality in ECE classrooms or programs are based on scores from instruments that involve conducting direct observations of ECE classrooms. Such instruments have been used to describe the quality of ECE large-scale studies, to examine relationships between the quality of ECE and children's developmental outcomes, to make comparisons of quality across different types of ECE classrooms and programs, to conduct external evaluations as part of QRIS, and as self-assessment tools. Many of the existing instruments designed to measure quality in ECE, however, were not initially developed for nor have they been validated for each of these purposes. Given such instruments are used to inform policy and practice in ECE, it is important to establish validity evidence for the myriad of purposes in which they are used.

Statement of the Problem

The heightened interest in conducting research to characterize dimensions of classroom or program quality and examine relationships to child outcomes has resulted in the need to identify whether existing instruments are valid for use for these intended purposes and can be used across different types of ECE classrooms or programs (Burchinal, Kainz, & Cai, 2011). Many existing instruments used to characterize quality in ECE include scales for which scores are based on observations conducted in classrooms or home care programs (Bryant, Burchinal, & Zaslow, 2011; Snow & Van Hemel, 2008). The instruments used widely to characterize dimensions of quality in ECE were developed theoretically, with substantial input from experts in ECE (Bryant et

al., 2011). Although there is empirical support for the reliability and validity of scores from these instruments, a need exists to gather additional validity evidence by using advanced methodologies to determine whether measures of quality differ across ECE classrooms that vary in the composition of children served (Bryant et al., 2011).

The Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms et al., 1998) and the Arnett Caregiver Interaction Scale (CIS; Arnett, 1989) have been used widely to characterize dimensions of quality in ECE classrooms or programs for different purposes. The ECERS-R (Harms et al., 1998) is a 43-item instrument with items organized under seven subscales. Each item on the ECERS-R has indicators of quality that range from inadequate quality to excellent quality. The ECERS-R has been used to assess the global, structural, and process quality of center-based preschool ECE classrooms. Scores on the ECERS-R are based primarily on indicators of proximal quality observed during classroom activities and routines. According to the developers, the ECERS-R provides information the extent to which three basic needs of children are met: (a) protection of health and safety, (b) social and emotional guidance and support, and intellectual stimulation (Environmental Rating Scale Institute [ERSI], n. d.). The developers of the ECERS-R proposed seven subscales for the instrument, each of which provides information about structural or process quality: (a) Space and Furnishings, (b) Personal Care Routines, (c) Language and Reasoning, (d) Activities, (e) Interaction, (f) Program Structure, and (g) Parents and Staff. The CIS (Arnett, 1989) is a 26-item instrument originally designed to characterize the quality of caregiver interactions in day care centers in a study conducted in Bermuda (Arnett, 1986; 1989). The CIS has subsequently been used to characterize process quality in other ECE

settings. Scores for the CIS are based on observations of proximal quality indicators related to caregiver interactions with children in center-based and home-based ECE programs and classrooms. In addition to providing information about the overall quality of caregiver interactions with children, there are four proposed subscales of the CIS: (a) Sensitivity, (b) Harshness, (c) Detachment, and (d) Permissiveness.

In the past 15 years, total, subscale, and derived composite scores from the ECERS-R have been used to make inferences about global, structural, and process quality in ECE programs and classrooms with differing compositions of children with and without special needs (e.g., Buysse et al., 1999; Grisham-Brown et al., 2010; Hestenes et al., 2008; La Paro et al., 1998). The composite scores were derived from principal components or factor analysis and hereinafter are referred to as derived composite scores. Some researchers have also used the ECERS-R to characterize the quality of ECE experienced by individual children with special needs enrolled in ECE classrooms (Clawsen & Luze, 2008; Knoche et al., 2006; Wall et al., 2006).

Total and subscale scores from the CIS have been used to characterize process quality for individual children with special needs enrolled in center-based ECE classrooms (Knoche et al., 2006; Wall et al., 2006). The extent to which it is appropriate to use these instruments, which were validated primarily in ECE programs for children without special needs (NSN), to characterize quality in preschool ECE classrooms in which children with special needs are enrolled (INC) has been called into question by researchers in the fields of early intervention and early childhood special education (Bailey et al., 1982; Buysse et al., 1999; Soukakou, 2012; Spiker, Hebbeler, & Barton, 2011).

Several validity studies have been conducted to gather evidence about the factor structures of the ECERS-R and the CIS (e.g., Cassidy, Hestenes, Hegde, et al., 2005; Clifford et al., 2005; Colwell et al., 2012; Gordon et al., 2013; Perlman et al., 2004; Sakai et al., 2003). Results from these studies have suggested factor structures that differ from those originally proposed by their developers. Despite empirical evidence from these and other studies about the factor structures of the ECER-S and CIS, inferences about quality or comparative quality of different types of ECE classrooms or programs continue to be made based on scores from the subscales originally proposed by the developers of the ECERS-R and CIS than on empirically derived composites or factor scores. In addition, none of the methodologies used to establish validity evidence for the ECERS-R and the CIS to date have tested the following three assumptions:

1. Conceptual equivalence of latent variables across groups (i.e., different types of classrooms),
2. Equivalent factor loadings of observed variables onto latent variables across groups, and
3. The extent to which observed variables are influenced to the same degree by measurement error (Vandenberg & Lance, 2000).

Violations of these assumptions significantly impact the interpretability of between-group comparisons of observed or derived composite scores such as between NSN and INC classrooms (Horn & McArdle, 1992). Such violations are also problematic when scores are aggregated across groups, because if scores that are not comparable are pooled, they cannot be used to make meaningful interpretations about the group as a whole.

Given the increasing number of studies in which scores from either the ECERS-R (Harms et al., 1998) or the CIS (Arnett, 1989) are aggregated across center-based ECE classrooms of different types (e.g., NSN, INC) or used as dependent measures to

examine differences in dimensions of quality across different types of classrooms, it is important to conduct appropriate statistical analyses to test the assumptions of measurement invariance across classrooms with different compositions of children with and without special needs.

Purpose of the Present Study

The purpose of the present study was to examine measurement invariance of the ECERS-R and the CIS across two types of center-based ECE classrooms using data from the ECLS-B: center-based inclusive preschool classrooms (INC) and center-based classrooms in which no children with special needs were enrolled (NSN). In the present study, an INC classroom was defined as a classroom in which: (a) children with and without special needs were enrolled at the time of ECLS-B data collection and (b) 75% or less of the children enrolled in the classroom were characterized as having special needs. A NSN classroom was defined as a classroom in which no children with special needs were enrolled at the time of ECLS-B data collection.

Multiple group confirmatory factor analysis was used to determine the extent to which there was measurement invariance for the ECERS-R and the CIS across these two types of classrooms. For each instrument, a confirmatory model was proposed based on a review of literature examining the structural validity of the instrument. An initial multiple group confirmatory model was estimated to determine the extent to which the instrument measured the same number of latent variables and whether the items defining these variables were the same across INC and NSN classroom groups (i.e., configural invariance, Horn, McArdle, & Mason, 1983). The primary purpose of the confirmatory model was to determine whether changes to the originally proposed measurement model were necessary. After the initial adequacy of the measurement

model was determined, a second multiple group confirmatory model (i.e., strict invariance model) was tested. The second confirmatory model determined the extent to which: (a) factor loadings, item threshold parameters, and residual variance parameters were invariant across the two types of classrooms and (b) there were differences in the means for the latent variables measured by the instrument across the two types of classrooms. In addition, regression analyses were conducted to examine the extent to which there were differences in derived composite scores across the two types of classrooms.

The aim of the present study was to examine the extent to which it is appropriate to make inferences about the quality of ECE in INC and NSN classroom groups based on derived composite scores from the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989) and, in the event evidence for measurement invariance was established, to examine group differences in latent variables measured by each instrument. Testing multiple group confirmatory factor models of measurement invariance provided information about how derived composite scores from these two instruments can be interpreted in INC and NSN classrooms. Examining group differences in empirically validated latent variables extended the literature related to quality across different types of ECE classrooms.

Conceptual Frameworks

Two conceptual frameworks were relevant for the present study. First, the conceptual framework that guided the conduct of the Early Childhood Longitudinal Birth Cohort (ECLS-B) study is described. The ECLS-B framework was relevant because the present study involved secondary analyses of data from a subsample of participants in the ECLS-B. Second, the analytical framework and measurement models related to

dimensions of quality applied to examine measurement invariance for the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989) are described.

Early Childhood Longitudinal Study (Birth Cohort) Conceptual Framework

The ELCS-B was a multiple-method, multiple-informant longitudinal study conducted with a nationally representative sample of children born in the United States in 2001. The purpose of the ECLS-B was to provide researchers, policymakers, families, and early childhood practitioners with information about the impact of children's early experiences on their health and development, with a particular emphasis on school readiness (Snow et al., 2007). The design for the ECLS-B study was based on the assumption that children's preparation for formal schooling begins at or before birth. Figure 1-1 depicts the conceptual framework that guided the design of the ECLS-B. This framework is rooted in ecological theory, which highlights the impact of multiple levels of influence on children's development (Bronfenbrenner & Crouter, 1983; Bronfenbrenner & Morris, 1998). These levels of influence include proximal influences (e.g., home, child care, early care and education) and distal influences (e.g., economic, social, education, legal, and political systems). As shown in Figure 1-1, the conceptual framework for the ELCS-B emphasized the interactions and relationships between multiple levels of influence thought to impact children's early development, including (a) child characteristics, (b) parent characteristics, (c) extended family/household characteristics, (d) community and schools, and (e) national and state policies (Snow et al., 2007).

Data related to each level of influence reflected in the conceptual framework were collected when the sample of children were 9 months old, 2 years old, 4 years old, and when they entered kindergarten. When children were 4 years old, the ECERS-R

(Harms et al., 1998) and the CIS (Arnett, 1989) were used to characterize dimensions of quality in the classrooms in which a subsample of children who participated in the study were enrolled. Quality variables were included in the ECLS-B study design because a number of studies have identified positive relationships between indicators of ECE classroom or program quality and children's short- and long-term developmental outcomes (e.g., Burchinal et al., 2008; NICHD ECCRN 1998, 2002, 2003; Peisner-Feinberg et al., 2000). The analytic sample for the present study was limited to children whose center-based preschool ECE classrooms were observed. Specifically, item-level data from the ECERS-R and the CIS were used to examine the measurement invariance of these instruments in classrooms defined as INC and NSN.

Measurement Models

A confirmatory factor analytic framework was applied in the present study to examine measurement invariance for the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989) across INC and NSN classrooms. Confirmatory factor analysis (CFA; Bollen, 1989) is a latent variable modeling technique used to test hypotheses regarding the interrelationships between observed variables that are presumed to measure one or more latent variables. CFA provides a mathematical framework in which researchers can validate conceptualizations of quality in ECE. Figure 1-2 depicts a conceptual model relating different dimensions of quality in ECE to child outcomes. Some proximal and distal quality indicators often associated with global, structural, or process quality are also shown in Figure 1-2. As shown in Table A-1, the items on the ECERS-R represent many of the indicators hypothesized to be associated with global, structural, and process quality (Cassidy, Hestenes, Hansen, et al., 2005; Cryer, 1999; Phillips &

Howes, 1987; Vandell & Wolfe, 2000). In addition, the items on the CIS are closely aligned with indicators of process quality (see Table A-2).

For the present study, multiple group confirmatory factor analyses (MG-CFA; Jöreskog, 1971) were conducted for each instrument of interest. MG-CFA is an extension of the CFA framework and permits the researcher to test hypotheses about the internal structure of an instrument under different conditions in which the instrument might be applied. MG-CFA was used to determine whether (a) the same number of latent variables were measured; (b) the items that defining the latent variables were equivalent across the two types of classrooms (i.e., configural invariance, Horn et al., 1983); (c) the factor loadings, threshold parameters, and residual variances were equivalent across the two types of classrooms (i.e., strict invariance; (Meredith, 1993); and (d) the means of latent variables measured were equal across the two types of classrooms. A review of existing literature was used to generate confirmatory factor models for the ECERS-R and the CIS.

ECERS-R measurement model

Figure A-1 displays the measurement model that was analyzed for the ECERS-R (Harms et al., 1998). The model was based on findings from a study in which exploratory factor analyses (EFA) were conducted using ECERS-R (Harms et al., 1998) data from the ECLS-B analytic sample (Gordon et al., 2013). Findings from this study provided empirical support for the three-factor model depicted in Figure A-1.

For the present study, 34 observed ECERS-R (Harms et al., 1998) item-level variables were hypothesized to be associated with three latent factors: (a) Activities and Materials, (b) Language and Interaction, and (c) Personal Care and Safety. Twenty items related to the arrangement of physical space for indoor and outdoor activities, the

provision of developmentally appropriate activities, and the provision of developmentally appropriate materials were selected as quality indicators for the first factor. The quality indicators for the second factor were nine items related to the supervision of children, provisions for children with special needs, adult-child interactions, and interactions among children. Five items related to personal care routines (e.g., meals, toileting), safety practices, and health practices were selected as quality indicators for the third factor.

The first two factors have been reported in a number of studies in which factor analysis or principal component analysis (PCA) of the ECERS-R (Harms et al., 1998) were conducted (Cassidy, Hestenes, Hegde, et al., 2005; Clifford et al., 2005; Sakai, Whitebook, Wishard, & Howes, 2003), with similar patterns of items to define them as those represented in the present model. In addition, the items associated with these factors are consistent with contemporary conceptualizations of structural and process quality provided by Cassidy, Hestenes, Hansen, et al. (2005). The third factor (i.e., Personal Care and Safety) is comprised of items that typically have been dropped from models as a result of previous studies (Cassidy, Hestenes, Hegde, et al., 2005; Clifford et al., 2005; Sakai et al., 2003), but that are considered important indicators of quality in ECE (Bryant et al., 2011; Cryer, 1999; Vandell & Wolfe, 2000). The measurement model proposed in the present study for the ECERS-R is consistent with the multidimensional conceptualization of quality in ECE represented in Figure 1-2 and includes many quality indicators identified as important to meet the basic needs of all children (ERSI, n. d.).

CIS measurement model

Figure A-2 depicts the bi-factor model that was used to examine measurement invariance of the CIS (Arnett, 1989). This model was based on information from a recent study in which CFA and item response theory models were used to examine the internal structure of the CIS with data from a similar analytic sample as the present study (Colwell, 2012). Colwell et al. (2012) examined traditionally structured models, in which each item was hypothesized to load onto only one factor, in addition to a bi-factor model. In the bi-factor model, all items loaded on a general factor (i.e., Caregiver Interactions) and an additional set of factors was proposed in which each item loaded on only one factor. The bi-factor model was used to account for the possibility that methodological factors are associated with positive or negative orientation of item wording rather than a substantive dimension of quality in ECE. When the positive or negative orientation of items was considered in the model, there was evidence to suggest the CIS is a unidimensional measure of caregiver interactions. As shown in Table A-2, all of the items on the CIS are related to interactions between the caregiver and child, making a unidimensional model of caregiver interactions defensible conceptually. Although some empirical studies have provided support for a multiple-factor model, these studies did not include procedures to control for item correlations that might have been due to methodological rather than substantive factors (Arnett, 1986, 1989; Whitebook et al., 1989). A bi-factor model was chosen for the present study because it is consistent with findings that were obtained through rigorous statistical analyses conducted with data from the ECLS-B.

ECLS-B Data Set

ECLS-B data used in the present study were obtained through a restricted-use license agreement with the Institute of Education Science (IES). The ECLS-B was a multiple-method, multiple-informant study funded by the U.S. Department of Education, National Center for Education Statistics (NCES) to provide detailed information about the early life experiences of a nationally representative sample of 14,000 children born in the United States in 2001. The ECLS-B data set includes data collected in five waves that occurred when the children were 9-months-old (2001), 2-years-old (2003), 4-years-old (preschool; 2005), 5- or 6-years-old (kindergarten; 2006), and 6- or 7-years-old (kindergarten; 2007). The final two waves of data collection are referred to as “kindergarten” waves. The goal of the ECLS-B was to follow children until they entered kindergarten, but not all children entered kindergarten in 2006. Children who did not enter kindergarten in 2006 were followed an extra year (i.e., the fifth data collection wave). Children who entered kindergarten in 2006 were not followed in the fifth data collection wave (Snow et al., 2007).

Data sources at each wave of data collection included parent interviews, parent questionnaires, and direct child assessments. Additional data sources for the 2- and 4-year data collection waves included interviews with each child’s ECE provider(s), interviews with center directors, and child care observations for a subsample of children who received non-parental care for more than 10 hours per week. Additional data sources for the kindergarten waves of data collection included data from teacher questionnaires and interviews with wrap-around ECE providers.

For the present study, secondary data analyses were conducted using data from the subsample of children whose center-based ECE classrooms were observed in the

4-year data collection wave and in which ECERS-R and CIS data were collected. Data from this subset of ECLS-B data were selected for the present study for several reasons. First, it includes data from multiple informants regarding children's health, development and educational experiences, which allowed for the identification of children with special needs, including children who had very low birth weight, identified disabilities, medical conditions diagnosed by a doctor, and developmental delays or emotional problems diagnosed by a professional. Second, the subsample of classrooms in which observations were conducted included a large sample of INC and NSN classrooms as defined in the present study. Third, the instruments used to characterize dimensions of quality in the ECLS-B were the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989). The ECERS-R and the CIS have been used widely to quantify dimensions of quality in a variety of ECE programs and classrooms, including the two types of classrooms of interest in the present study. Finally, item-level data from the ECERS-R and the CIS were included in the ECLS-B data file to allow researchers to conduct validity studies, including factor analyses (Snow et al., 2007).

The large sample size and the diversity of children and ECE classrooms observed as part of the ECLS-B provided an opportunity to generate validity evidence about the internal structure of these instruments and to examine various aspects of measurement invariance across INC and NSN classrooms. Evidence of measurement invariance across these two types of classrooms has not yet been established for either instrument. The present study adds to the extant literature regarding internal structure validity evidence for the ECERS-R and CIS. Specifically, the present study provides

information about the extent to which derived composite scores from these instruments are comparable across INC and NSN classrooms.

Research Questions

The following research questions guided the secondary analyses conducted to examine aspects of measurement invariance of the ECERS-R (Harms et al., 1998) and CIS (Arnett, 1989) across INC and NSN classrooms:

1. Are the same number of latent variables measured by the ECERS-R in INC and NSN classrooms? Are the same number of latent variables measured by the CIS in INC and NSN classrooms?
2. To what extent are the items that define the latent variables measured by the ECERS-R the same in INC and NSN classrooms? To what extent are the items that define the latent variables measured by the CIS the same in INC and NSN classrooms?
3. To what extent are the factor loadings, item threshold parameters, and error variance parameters for items on the ECERS-R invariant across INC and NSN classrooms? To what extent are the factor loadings, item threshold parameters, and error variance parameters for items on the CIS invariant across INC and NSN classrooms?
4. To what extent are means of the latent variables measured by the ECERS-R different between INC and NSN classrooms? To what extent are means of the latent variables measured by the CIS different between INC and NSN classrooms?

Significance of the Study

The present study extends current knowledge about the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989) in a number of ways. First, findings from the present study replicated previous evidence for the structural validity of the ECERS-R and CIS using a large sample from a nationally representative study of young children. Second, the present study extends previous research conducted to establish validity evidence for these two instruments based on internal structure. To date, there have been no published studies identified that have examined the measurement invariance of these

instruments across INC and NSN classrooms. Both the ECERS-R and the CIS have been used in research and program evaluation endeavors that have had implications for early childhood practice and policy. Findings from the present study provide information about whether the latent variables measured by the ECERS-R and CIS are equivalent across INC and NSN classrooms. Findings from the present study also provide information about whether derived composite scores obtained from the ECERS-R and CIS are equivalent across INC and NSN classrooms. The types of validity evidence examined in the present study are essential to ensure accurate interpretation of findings from comparisons of quality across these types of ECE classrooms and from data that have been aggregated across these two types of classrooms. In addition, the present study is one of only a few studies identified in which comparisons of quality across INC and NSN classrooms were made using empirically validated latent variables.

Findings from this study will help researchers, policymakers, and practitioners make informed decisions about future use of ECERS-R and CIS to characterize dimensions of quality, the interpretation of findings from previous studies, and the need for future research examining the measurement invariance of these instruments in programs and classrooms other than those described in the present study. In addition, findings from the present study will extend the current knowledge base regarding differences in quality across two types of ECE center-based classrooms.

Delimitations

The present study involved secondary analysis of data collected via observations in ECE classrooms from a subset of data contained in the restricted-use ECLS-B data file. The present study did not involve primary research design or data collection. The data used in the present study were cross-sectional and drawn from a subsample of the

ECLS-B participant pool that was eligible to have observations of their ECE classrooms in the preschool wave of data collection. Although similar data were available for the 2-year follow up of the ECLS-B, those data were not included in the secondary analyses conducted for the present study, because the instruments used to characterize the quality in ECE classrooms in the 2-year data collection wave differed slightly than those used in the preschool wave of data collection.

The primary questions that guided the secondary analyses in the present study addressed the issue of measurement invariance for two instruments that were used during observations of center-based ECE classrooms in which preschool-aged children participating in the ECLS-B were enrolled. The ECERS-R (Harms et al., 1998) was used to characterize global quality of these classrooms and the CIS (Arnett, 1989) was used to characterize the quality of caregiver interactions. Data used in the present study were drawn from observations of center-based preschool ECE classrooms only. Although they were available in the data file, ECERS-R and CIS data from observations of home-based preschool ECE classrooms were not included in the present analyses, because the instrument used to assess global quality in these classrooms differed from the instrument used to assess global quality in center-based ECE classrooms.

The focus of the present study was measurement invariance of the ECERS-R (Harms et al., 1998) and CIS (Arnett, 1989) across INC and NSN classrooms. Measurement invariance was not examined for any other instruments, ECE classrooms, or subgroups of children.

Limitations

A number of limitations of the present study are noted. First, data used in the present study are not from a nationally representative sample of classrooms. The

purpose of the ECLS-B was to provide detailed information about health, development, and early life experiences of a nationally representative sample of children born in the United States in 2001. The instruments of interest for the present study provided information about classrooms in which these children were enrolled. The ECLS-B data set provides a large analytic sample of item-level data for these instruments, which is optimal for conducting the analyses carried out in the present study. Because the sampling unit for the ECLS-B was children and not classrooms, it cannot be inferred that the analytic sample used in the present study is nationally representative of center-based preschool ECE classrooms. In addition, the subsample of children whose ECE classrooms were observed in the preschool wave of data collection excluded some subgroups of children who might have been enrolled in center-based ECE classrooms at the time of data collection but were not sampled due to limitations in resources. Although ECLS-B data were weighted to account for sampling procedures and nonresponse rates at the time of data collection and across data collection waves, the weighting procedures did not account for the exclusion of classrooms that were not sampled due to limitations in resources.

Second, there was not a complete item pool from the ECERS-R (Harms et al., 1998). Six items pertaining to provisions for staff and families were not administered as part of the ECLS-B observation protocol, because they could not be observed directly or because the information could be obtained in another way (Snow et al., 2007). Three additional items were dropped from the item pool for the secondary analyses conducted in the present study due to missing data related to applicability of the item to the classroom in which the instrument was administered (i.e., nap/rest, use of

TV/computers, provisions for children with disabilities). Previous studies that have examined the internal structure of the ECERS-R have generally yielded factor structures for which these items have not had salient factor loadings (e.g., Cassidy, Hestenes, Hegde, et al., 2005) or they have been conducted with an item pool that is similar to the item pool used in the present study (e.g., Clifford et al., 2005; Gordon et al., 2013). Inclusion of these items could result in a different factor structure than the one modeled in the present study.

Third, the version of the CIS (Arnett, 1989) used in the ECLS-B was adapted slightly from its original version and from versions used in other large-scale studies (Snow et al., 2007). In general, the adaptations to the items were the addition of examples to facilitate accurate scoring. Potential systematic differences in scoring of the CIS for the ECLS-B compared to other studies in which it has been used might limit the extent to which findings from the present study can be generalized to other situations.

Fourth, item and latent variable scores for the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989) reflect the quality of an ECE classroom rather than the quality of ECE experienced by a particular child. Although these scores likely reflect the highest level of quality an individual child could receive, it is possible that a particular child's experience could be lower than that reflected in item or latent variable scores on the ECERS-R or the CIS. Classifications of classrooms as either INC or NSN were made based on the characteristics of a classroom and not a program at the time of data collection. It is not known if children with special needs were enrolled in the observed classroom at any point prior to or after ECLS-B data collection or whether the program

in which the observed classroom was situated had other classrooms that included children with special needs. The comparisons of latent variables made in the present study reflect differences in the quality in ECE across INC and NSN classrooms. They do not represent comparisons of quality for groups of individual children who participated in the ECLS-B.

A number of analytic decisions were made in order to conduct secondary analyses using the ECLS-B data, including recoding data to generate composite variables, selection of weights, and the selection of analytic techniques. Descriptions of these decisions are provided in Chapter 3 and should be considered when interpreting findings from the present study.

Chapter Summary

The purpose of this chapter was to provide an introduction to the present study. Background information was presented to (a) describe contextual factors that have informed interest in and characterization of dimensions of quality in ECE programs and classrooms; (b) describe how conceptualizations of quality in ECE have changed over time; and (c) provide an overview of the ways in which global, structural, and process quality have been studied in ECE programs and classrooms. A statement of the problem highlighted the need for technically adequate instruments to measure dimensions of quality in ECE programs and classrooms, with an emphasis on the importance of examining the measurement invariance of two instruments that have been used widely in different types of ECE classrooms or programs. The purpose of the study was presented, followed by a description of the conceptual frameworks used. The research questions that guided the secondary analyses conducted for the present

study were presented, followed by the significance of the study. Delimitations and limitations were also presented.

Table 1-1. Percentage of Children Under Age 5 Attending Center-Based ECE Programs in 2011 by Employment Status and Characteristics of Mother

Characteristic of Mother	Mother Employed		Mother Unemployed	
	Total No. of Children (in thousands)	% in Care	Total No. of Children (in thousands)	% in Care
Race/Hispanic Origin				
White	7724	34.7	6692	10.3
Black	1462	40.6	4429	21.7
Asian	471	33.4	1420	23
Hispanic	2592	24.3	403	7.8
Marital Status				
Married	7068	35.7	5946	10.8
Separated, divorced, widowed	744	37.5	606	17.1
Never married	2260	35.7	2428	17.9
Poverty Status				
Below poverty level	1599	26.9	3076	14.8
At or above poverty level	8320	37.9	5339	12.1

Note. Source: "Who's Minding the Children? Child Care Arrangements: Spring 2011" by L. Laughlin, 2013.

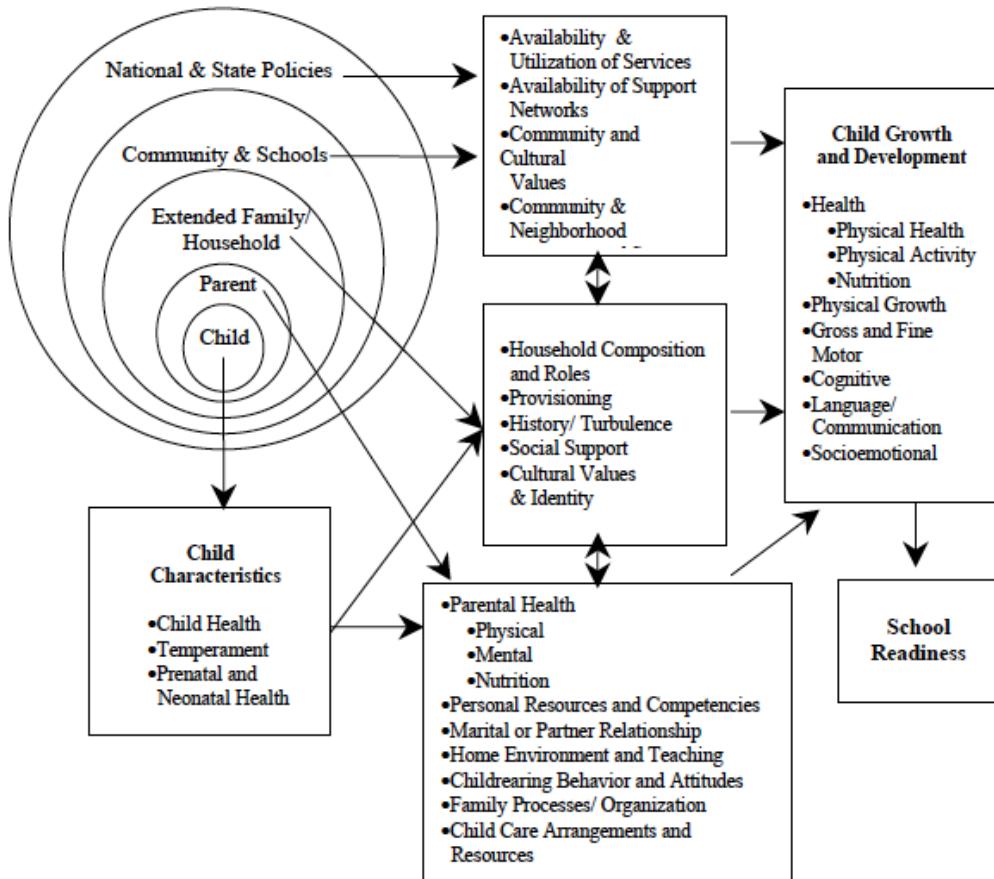


Figure 1-1. ECLS-B Conceptual Framework. Reprinted with permission from “Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), Preschool Year Data File User’s Manual (2005-2006)” by K. Snow et al., 2007, p. 6.

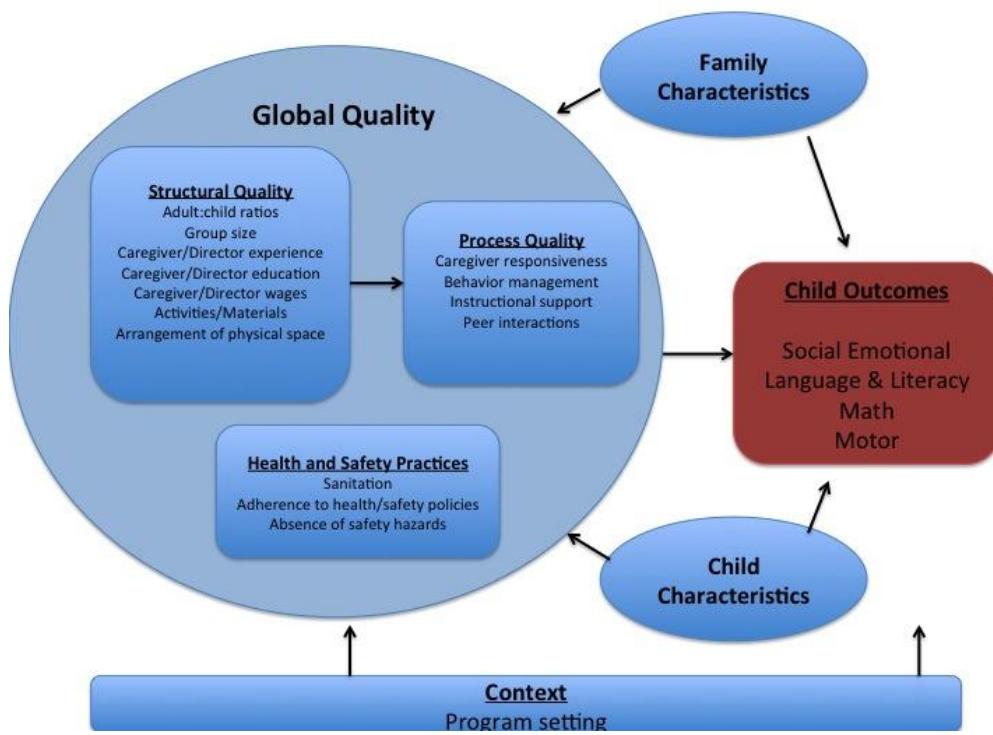


Figure 1-2. Conceptual Model of Relationships Between Global, Structural, and Process Quality and Child Outcomes.

CHAPTER 2

REVIEW OF THE LITERATURE

This chapter describes a review of the literature relevant to the present study.

The purpose of the literature review is to provide background information and a rationale for the present study. The literature review covers four major topics: (a) defining quality in ECE, (b) measuring quality in ECE classrooms and programs, (c) Early Childhood Environment Rating Scale-Revised (Harms et al., 1998), and (d) Arnett Caregiver Interaction Scale (Arnett, 1989). The first topic provides background information about how quality in ECE has been conceptualized by researchers. The second topic focuses on the ways in which researchers have assessed the indicators of quality experienced by children that have the most direct influences on children's health and development. The second topic also provides a rationale for the study by highlighting the need to establish evidence of measurement invariance for measures of quality used in a variety of ECE classrooms or programs. This topic includes a review of empirical studies that have examined measurement invariance for instruments used in ECE. The third and fourth topics provide information about the instruments for which measurement invariance was examined in the present study. For each instrument, the following information is provided: (a) development of the instrument; (b) a review of empirical studies examining the factor structure of the instrument; and (c) a review of empirical studies in which the instrument was used to measure quality in center-based ECE preschool classrooms where children with special needs were enrolled.

Search Procedures

Research studies, articles, and reports reviewed for this study were identified from a number of sources. First, peer-reviewed articles and technical reports were

identified using the following electronic databases: Educational Resource Information Center (ERIC) and EBSCO Host Platform for Academic Search Premier, Education Full Text, Psychology and Behavior Sciences Collection, Professional Development Collection, PsychInfo, and Social Sciences Full Text. All studies, articles, and reports identified from the searches of these databases were screened to determine whether they included information relevant to any of the four topics addressed in the present literature review. A final step of the literature search included a hand-search of reference lists for all studies, articles, and reports identified from electronic searches. Resources identified as part of the hand search also underwent screening procedures to determine their relevance for the present study. The search procedures used to identify and screen sources for the present literature review are described, followed by a summary of the number of sources located, screened, and retained.

Search Procedures for Topics 1 and 2

Each of the following sets of keywords was entered separately and then in combination with each other as a first step to search for articles and technical reports related to defining and measuring quality in ECE: educational quality, classroom environment, day care effects, early childhood, preschool, child care quality, day care quality, and experiences. The purpose of beginning with these terms was to locate as many sources related to quality in ECE as possible. Within each of these initial searches the following additional sets of keywords were entered in combination with the phrases “global quality”, “process quality”, “structural quality”: teacher-child relationships, definition, measure, and measurement. For this portion of the literature review, only articles or technical reports that summarized definitions or conceptualizations of quality in ECE from previous research, described conceptual

models for defining quality in ECE, included discussions of issues related to measuring quality in ECE were reviewed. Articles and reports focused on the measurement properties of specific instruments designed to measure quality in ECE were not reviewed under topics 1 and 2. In addition to articles and reports from these searches, several books and book chapters pertaining to the conceptualization of quality in ECE, measurement of quality in ECE, and examining reliability and validity evidence for instruments used in educational settings were used to guide the discussion of topics 1 and 2.

In addition to presenting general information about the measurement of quality in ECE, a review of empirical research studies in which measurement invariance was examined for instruments designed for use in preschool programs is included under topic 2. To identify empirical studies in for this section of the literature review, the following keywords were used in combination with the terms “preschool,” “young children,” and “early childhood”: measurement invariance, factorial invariance, factor invariance, multiple group factor analysis, mutli-group factor analysis. Studies identified using these search terms were screened and were only retained for review if they met the following criteria: (a) multiple group factor analysis was use to examine measurement invariance; and (b) the instrument of interest in the study was designed for use with preschool-age children, families of preschool children, or in programs where ECE was provided to preschool children. Studies meeting these two criteria were also reviewed to identify whether any of the studies focused on an instrument used to measure quality in ECE.

Search Procedures for Topics 3 and 4

The user's manual for the ECERS-R (Harms et al., 1998) was the primary source of information for the third topic addressed in the present literature review. In addition, the instrument name was entered as a search term in the databases listed previously. To identify empirical studies examining the factor structure of the ECERS (Harms & Clifford, 1980) or ECERS-R (Harms et al., 1998), the following search terms were each entered in combination with the instrument name: measure, principal component analysis, factor analysis, and factor. Articles, studies, and reports included under this topic in the literature review met the following criteria: (a) principal components analysis (PCA), exploratory factor analysis (EFA), confirmatory factor analysis (CFA), or multiple group factor analysis (MG-FA) was conducted; (b) the measure of interest was the English version of either the ECERS (Harms & Clifford, 1980) or the ECERS-R (Harms et al., 1998); (c) the authors provided information about the methodology used to conduct the analyses; and (d) the authors' description of the internal structure of the scale included, at a minimum, an explanation of the pattern of item loadings onto identified factors. The latter two criteria were applied because a number of studies briefly mentioned factors identified from a factor analysis, but they did not provide sufficient information to conduct a critical analysis of the methodology or findings from such examinations.

Additional keywords used to identify empirical studies in which the ECERS or ECERS-R was used to characterize the quality in ECE in center-based preschool classrooms where children with special needs were enrolled. Each of the following search terms was entered in combination with the instrument name: disabil*, special needs, special education, and early intervention. Studies identified using these

keywords were screened for relevance to the present study and were only retained for if they met the following criteria: (a) data were collected in preschool classrooms; (b) the sample included children with special needs or classrooms that provided ECE to these children; (c) either the ECERS or ECERS-R was used to characterize global, structural, or process quality. The same search procedures and article screening procedures were used to identify similar types of articles, studies, and technical reports related to the CIS (Arnett, 1989).

Summary of Sources Reviewed

A total of 259 articles and technical reports were located and screened for their relevance to one of the four topics presented in the present review of the literature as a result of the electronic and hand searches conducted. Twenty-five empirical studies focused on measurement invariance of an instrument designed for use in preschool ECE settings, however, only one study focused on measurement invariance for an instrument designed to measure quality in ECE. Eight empirical studies were identified that met the inclusion criteria for empirical studies examining the factor structure of the ECERS (Harms & Clifford, 1980) or the ECERS-R (Harms et al., 1998). Two empirical studies met the inclusion criteria for empirical studies examining the factor structure of the CIS (1989). A total of eight empirical studies met the inclusion criteria for studies using the ECERS, ECERS-R, or CIS to characterize the quality in ECE in ECE classrooms where children with special needs were enrolled. Three of these studies used the ECERS to characterize the quality in ECE; three studies used the ECERS-R; and two studies used both the ECERS-R and the CIS. Information from the sources reviewed is synthesized and presented in the following sections of this chapter. Each of the four topics noted previously is presented separately.

Defining Quality in ECE

The purpose of this section of the literature review is to describe the ways in which conceptualizations about quality in ECE have evolved. Quality can be defined from multiple perspectives. Four perspectives described are (a) the top-down perspective, which focuses on components of quality in ECE that have been described by researchers; (b) the inside-out perspective, which focuses on features of quality defined by caregivers and directors; (c) the bottom-up perspective, which encompasses quality in ECE from the child's perspective; and (d) the outside-in perspective, which is defined by parents' perceptions of quality (Ceglowski, 2004; Ceglowski & Bacigalupa, 2002; Katz, 1995; Harrist, Thompson, & Norris, 2007). Due to the extensive use of the instruments of interest in the present study by researchers, the present literature review focused on the top-down conceptualization of quality in ECE.

Since the first large-scale studies were conducted to examine global, structural, and process quality in ECE programs in the United States (e.g., Helburn, 1995; Ruopp et al., 1979; Whitebook et al., 1989), improving quality in ECE classrooms and programs has become a national priority. In addition, a major focus of early childhood research is to examine dimensions of quality in ECE and to examine relationships between quality or thresholds of quality and child developmental and learning. Over the past 3 decades, the same general definition of quality has guided early childhood research, practice, and policy; however, this definition lacks specificity and has been conceptualized in slightly different ways by different researchers (La Paro et al., 2012). In general, there has been consensus that quality in ECE is defined by the following six core components:

1. Safety of care, which includes adequate supervision of children and the safety of physical surroundings, such as toys, equipment, and physical space;
2. Health practices, which includes opportunities for physical activity, rest, maintenance of sanitary conditions for children, and meeting children's nutritional needs;
3. Implementation of developmentally appropriate practices, which include opportunities for children to direct their own engagement in a variety of activities and instructional support to promote children's learning and development within these activities;
4. Positive interactions with adults;
5. Promoting emotional growth, which includes practices that support children's abilities to operate independently, cooperatively, securely, and competently; and
6. Promoting positive relationships with other children (Cryer, 1999).

These components are evident in the guidelines for developmentally appropriate practice in ECE (Bredekamp & Copple, 2009), recommended practices in early intervention and early childhood special education (Sandall, et al., 2005), and observable quality indicators included in measures used to quantify quality in ECE classrooms and programs (e.g., Arnett, 1989; Harms et al., 1998; Pianta, La Paro, & Hamre, 2008). Variables that have been conceptualized as indicators of the core components of quality have been described as either proximal or distal quality indicators (Dunn, 1993). Proximal quality indicators have a direct influence on the children's health, safety, and development because they are experienced directly by children (e.g., adult-child interactions, materials, activities). Distal quality indicators have an indirect influence on children's health, development, and safety (Dunn, 1993; Love et al., 1996).

A lack of consensus about which proximal and distal quality indicators are most important and what dimensions of quality are associated with these indicators has resulted in a lack of specificity in the definition constructs of quality in ECE (Layzer &

Goodson, 2006, La Paro et al., 2012). Two key dimensions of quality that are commonly referred to are process quality and structural quality. Process quality has been defined broadly as the quality of dynamic processes that occur in the ECE program or classroom, whereas structural quality has been described as the dimension of quality that comprises the framework in which dynamic processes operate (Cryer, 1999; Phillips & Howes, 1987; Love et al., 1996; Vandell & Wolfe, 2000). Global quality has been conceptualized as the overall quality of a program and is usually characterized by both process and structural quality, as well as indicators of quality related to health and safety practices (Cryer, 1999; Love, et al., 1996; Phillips & Howes, 1987; Vandell & Wolfe, 2000). The broad definitions of global, structural, and process quality have been accepted widely among researchers; however, the organization of proximal and distal quality indicators under these dimensions varies widely.

In most early conceptualizations of quality in ECE, process quality was defined by characteristics of a program or classroom actually experienced by children (Cryer, 1999, Love et al., 1996; Phillips & Howes, 1987; Vandell & Wolfe, 2000). Using this definition, indicators of process quality included proximal indicators such as activities, materials, physical space, teacher-child interactions, child-child interactions, safety practices, health practices, and personal care routines. In a review of quality indicators used to characterize quality on the ECERS-R (Harms et al., 1998), Cassidy, Hestenes, Hansen, et al. (2005) challenged this conceptualization of process quality by asserting that indicators of process quality should be limited to those that require “human interactions among individuals” (p. 510). Cassidy, Hestenes, Hansen, et al. reasoned that in order for a dynamic process to occur during activities or in the presence of

classroom materials, an adult must be actively involved with children as they interact with the physical environment. Under this definition, proximal quality indicators related to features of the physical environment, such as room arrangement, the provision of materials, and the provision of activities are characterized as indicators of structural rather than process quality. Although there has been some empirical evidence to suggest such aspects of the physical environment are associated with a different dimension of quality than indicators of quality requiring some type of interaction between individuals (Cassidy, Hestenes, Hegde, et al., 2005; Clifford et al., 2005; Sakai et al., 2003; Whitebook et al., 1989), the notion that aspects of the physical environment are indicators of structural rather than process quality has not been adopted widely.

Conceptualizations of structural quality have also evolved over time. In early conceptualizations, structural quality represented “regulatable” features of ECE programs (Phillips & Howes, 1987). Under this definition, only distal quality indicators, such as adult-child ratios, group size, and teacher and director qualifications, were conceptualized as indicators of structural quality (Cryer, 1999; Phillips & Howes, 1987). The definition of structural quality was then expanded to include other distal quality indicators, such as staff wages, staff turnover, parent fees, and parent involvement (Cryer, 1999; Love et al., 1996; Vandell & Wolfe, 2000). Following this expansion in the conceptualization of structural quality, indicators of structural quality were generally perceived as inputs to process quality (Cryer, 1999). Cassidy, Hestenes, Hansen, et al. (2005) proposed structural quality is defined by characteristics of the ECE program or classroom that are “independent of human interaction between individuals” (p. 511). Under this definition, both proximal and distal quality indicators can be conceptualized

as indicators of structural quality. Proximal quality indicators that are characterized as indicators of structural quality under this definition include the arrangement of physical space, the presence of developmentally appropriate materials, and opportunities to engage in developmentally appropriate activities. The definition of structural quality proposed by Cassidy, Hestenes, Hansen, et al. is aligned with the broader definition of structural quality (i.e., inputs to process quality), because these features of the physical environment are often antecedents that set the occasion for more dynamic processes to occur between individuals, but their presence alone does not guarantee that these processes will occur.

For the present study, the definitions of process and structural quality proposed by Cassidy, Hestenes, Hansen, et al. (2005) were adopted to aid in the interpretation of the internal structure of the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989), because they provide more operational definitions with which to interpret specific indicators of quality in ECE than the broader definitions of structural and process quality that have been adopted previously. Furthermore, the specific conceptualizations of structural and process quality provided by these authors are aligned with the broad definitions characterizing process quality as representative of dynamic processes in the classroom and structural quality as representative of inputs to process quality.

Measuring Quality in ECE

The growing interest nationwide in characterizing and evaluating quality in ECE has necessitated the development of instruments that can be used reliably to make valid interpretations about different dimensions of quality. In general, information about distal quality indicators has been collected via interviews or self-reports from ECE providers and directors (Layzer & Goodson, 2006; Snow & Van Hemel, 2008). Proximal

quality indicators are observable; therefore, a number of observational methods have been employed by researchers, policymakers, and practitioners to quantify proximal quality indicators. The measures of interest in the present study were developed to provide information about proximal quality indicators that have been associated with different dimensions of quality in ECE.

The focus of this section of the literature review is issues related to measuring proximal quality indicators, because they are most relevant to the present study. A general description of the approaches used to quantify proximal quality indicators is presented, followed by approaches for establishing validity evidence for instruments designed to measure different dimensions of quality in ECE programs and classrooms. This portion of the literature review concludes with a review of empirical studies examining the measurement invariance of instruments designed for use in ECE programs.

Approaches for Quantifying Quality in ECE

Information about proximal quality indicators is best obtained through observations of ECE classrooms, because these indicators are experienced directly by children during classroom activities and routines. Proximal quality indicators include caregiver sensitivity and responsiveness, instructional supports provided to children, support for social skills, behavior management, the provision of activities and materials to support children's learning in a variety of domains (e.g., literacy, math, science, language), adaptations for children with special needs, arrangement of the physical environment, and health and safety practices (Snow & Van Hemel, 2008). Table 2-1 displays several observational instruments used widely to assess proximal quality indicators in ECE and the purposes for which they have been used. As shown in Table

2-1, three major observation-based approaches have been utilized to quantify proximal quality indicators: (a) observational coding systems, (b) judgment-based observational rating scales, and observational checklists (Snow & Van Hemel, 2008; Halle et al., 2010). Each of these approaches allows the user to obtain a different level of information regarding proximal quality indicators in ECE classrooms.

Observational coding systems are generally designed to provide micro-level information about specific indicators of quality in ECE, such as caregiver sensitivity or children's exposure to instructional supports (e.g., NICHD ECCRN, 2000; Ritchie, Howes, Kraft-Syre, & Weiser, 2001). Data regarding these indicators are generally collected through time-sampling techniques that allow the researcher to quantify the frequency or quality of specific behaviors, which have been operationalized and manualized for the user. Typically, observational coding systems require extensive training for users to administer reliably (Harms & Clifford, 1982) and they only provide information about very specific quality indicators. They are relatively labor and time intensive, but can yield important information about specific dimensions of quality at a micro level. Observational coding systems are not used as often as other observation-based instruments designed to assess proximal quality indicators in ECE classrooms (Harms & Clifford, 1982; Layzer & Goodson, 2006).

Judgment-based observational rating scales are another type of observation-based instrument used to quantify proximal quality indicators in ECE. Judgment-based, observational rating scales allow users to apply ratings that apply to process quality indicators on the basis of observations of classroom activities and routines. Typically, indicator scoring is based on 2 to 4 hours of observation (Snow & Van Hemel, 2008).

For example, observers watch actions and behaviors occurring in the classroom, assess the variety and types of materials present in the room, how the room is arranged, or provisions that are made for young children with special needs in order to make observation-informed judgments about a rating scale score to apply. As shown in Table 2-1, the content of judgment-based rating scales used to assess different dimensions of quality in ECE varies widely. Some rating scales provide information about global, structural, and process quality (e.g., Harms et al., 1998; High Scope, 2003), whereas other rating scales provide information about specific dimensions of quality (e.g., Arnett, 1989; Hyson, Hirsh-Pasek, & Rescorla, 1990; ; Pianta et al., 2008; Smith et al., 2002; Stipek & Byler, 2004; Sylva, Siraj-Blatchford, & Taggart, 2003).

Many judgment-based observational rating scales provide total scores that summarize ratings across all of the items on the scale, in addition to subscale scores that summarize ratings across items that are hypothesized to represent a particular dimension or construct related to quality (Bryant et al., 2011; Vandell & Wolfe, 2000).

As shown in Table 2-1, judgment-based observational rating scales are the most common type of instrument used to assess quality in ECE. One potential reason for this is they provide relatively detailed information about proximal quality indicators, and the response-cost for training observers to reliability standards is less than that for observational coding systems (Harms & Clifford, 1982).

A third approach for quantifying proximal quality indicators in ECE is the use of observational checklists. Observational checklists are often used to provide information about the presence or absence of proximal quality indicators (Abbot-Shim & Sibley, 1992; Abt. Associate, Inc., 2006, Smith et al., 2002). As with judgment-based rating

scales, these instruments can yield global or subscale scores. Observers can be trained to use these instruments reliably with relatively limited training; however, the information provided by these instruments is sometimes limited compared to observational coding systems or to judgment-based observational rating scales, because they do not offer a continuum of response scores or the ability to characterize the frequency or quality of specific behaviors. Observation checklists can, however, be useful for providing information about the materials and activities available to children.

Content and Scope of Quality Measures in ECE

Each of the approaches for quantifying proximal quality indicators can be used to collect information from various domains of content. Layzer and Goodson (2006) described two categories of instruments designed to measure proximal quality indicators in ECE: measures of caregiving process and global measures of quality. Measures of caregiving process are designed to capture the dynamic aspects of quality and tend to provide in-depth information about a particular aspect of the caregiving process. Typical content foci of such instruments include instructional supports provided to children during classroom activities and routines (e.g., Ritchie, et al., 2001) and the nature of interactions that occur among children and adults in the classroom (e.g., NICHD ECCRN, 2000). Global measures of quality are used to synthesize information across multiple quality dimensions (Vandell & Wolfe, 2000). As with measures of caregiving process, global measures of quality can vary in their content focus. For example, some global measures of quality were designed to synthesize information about quality across a broad range of quality indicators, such as the provision of materials and activities, interactions among individuals in the classroom, and health and safety practices (e.g., Harms et al., 1998). Other global measures of

quality were designed with less emphasis on health and safety provisions and more emphasis on proximal quality indicators related to classroom climate and instructional supports (e.g., Pianta et al., 2008; Stipek & Byler, 2004) or a particular instructional domain, such as early literacy (e.g., Smith et al., 2002) or mathematics (National Institute for Education Research, 2007).

Most of the observational instruments currently available to quantify proximal quality indicators provide information based on observations in ECE classrooms. One criticism of these instruments is they do not provide information about individual children's experiences (Clawson & Luze, 2008; Layzer & Goodson, 2006; Snow & Van Hemel, 2008; Spiker et al., 2011), despite having been designed to capture proximal quality indicators that, by definition, are experienced directly by children. Most of these instruments provide a fairly detailed 'snapshot' of the quality of the social environments in ECE classrooms. The extent to which they provide information about the quality of learning environments in ECE classrooms is generally limited to the types of materials and activities provided to children (Snow & Van Hemel, 2008).

Establishing Validity Evidence for Measures of Quality in ECE

Limitations in the method for collecting information, the content focus, and the scope of and purpose for using the instrument should be taken under careful consideration when choosing an instrument with which to measure global, structural, or process quality in ECE programs or classrooms. All of the instruments listed in Table 2-1 were originally developed and validated for administration under specific circumstances established by the instrument developers (Snow & Van Hemel, 2008). Regardless of the intent for which they were originally developed, many of these instruments have been used for numerous purposes by different observers. ECE

program directors and staff have used these instruments as self-assessment tools to guide quality improvement and professional development endeavors. Researchers have used a variety of observational instruments to provide descriptive information about large, heterogeneous samples of ECE classrooms or programs and to make judgments about the quality in ECE programs nationwide (e.g., Helburn, 1995; Layzer et al., 1993; Whitebook et al., 1989). Scores from these instruments have also been used as dependent variables to compare ECE quality across different types of ECE programs or classrooms, to measure the efficacy of early childhood interventions, and to predict child development (La Paro, et al., 2012). Finally, several of the instruments listed in Table 2-1 have been used by external evaluators for accountability purposes or to assign quality ratings to ECE programs as part of statewide QRIS (Bryant et al., 2011; Snow & Van Hemel, 2008).

The extent to which these instruments have been used appropriately to make inferences about quality in ECE is dependent on the validity evidence for interpreting the scores yielded in circumstances under which the instruments were administered. The *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) described five sources of validity evidence: (a) evidence based on content, (b) evidence based on response processes, (c) evidence based on internal structure, (d) evidence based on relations to other variables, and (e) evidence base on consequences of testing. Evidence based on content refers to the extent to which the content of an instrument represents the construct(s) it is hypothesized to measure. In the case of observational instruments, evidence based on

response processes refers to evidence to suggest that observers apply ratings as intended, without influence from factors that are irrelevant to the intended interpretation of instrument scores. Evidence based on internal structure refers to the extent to which items on an instrument measure the constructs for which score interpretations are made. Evidence based on relationships to other variables refers to correlations between a measure of interest and criterion variables that are hypothesized to be associated with the constructs measured by the instrument. Evidence based on consequences of testing refers to whether or not the intended or unintended consequences that arise from the interpretation of instrument scores are aligned with the intended use of the instrument (Snow and Van Hemel, 2008; Snyder, McLean, & Bailey, in press).

The focus of the present study was to provide evidence related to internal structure for the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989). Although there have been some empirical studies examining the factor structure of these instruments, they studies typically have been limited to exploratory analyses designed to examine the extent to which the instrument items measure the constructs of quality proposed by the instrument developers.

A different form of evidence based on internal structure involves the establishment of measurement invariance for the instrument across different populations or groups. Measurement invariance refers to the extent to which instruments yield measures of the same constructs under different conditions (Horn & McArdle, 1992). When measurement instruments are administered under multiple conditions without evidence of measurement invariance across these conditions,

interpretations of instrument scores are “at best ambiguous and at worst erroneous,” (Steenkamp & Baumgartner, 1998, p. 78). One “condition” for which measurement invariance is often examined is different populations or subgroups of respondents.

Given the heterogeneity of the ECE programs and classrooms in which measures of quality in ECE are applied, this form of validity evidence is particularly important to establish.

Latent variable models typically used to establish validity evidence based on internal structure (e.g., item response models, confirmatory factor analyses, exploratory factor analyses) can be used to examine whether the internal structure of a particular instrument is consistent across groups of interest (Hui & Triandis, 1985; Jöreskog, 1971; Muthén & Muthén, 2008-2012; Vandenberg & Lance, 2000). Such models allow researchers to examine multiple levels of measurement invariance.

The most fundamental level of measurement invariance is configural invariance. Configural invariance refers to the extent to which an instrument measures the same latent variables across the groups of interest (Horn et al., 1983). Weak invariance refers to the extent to which factor loadings between observed and latent variables are similar across the groups of interest (Rock et al., 1978). Strong invariance refers to the extent to which item score means are invariant across groups (Meredith, 1993). Strict invariance invariance refers to the extent to which residual variances are the same across groups (Meredith, 1993). Evidence of strict invariance suggests latent variable scores are comparable and equally accurate across groups (Steenkamp & Baumgartner, 1998). In the present study, latent variable modeling techniques were

used to examine the extent to which there was evidence of measurement invariance for the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989).

Empirical Studies Examining Measurement Invariance in ECE Programs

Measures of quality in ECE have been administered for numerous purposes in a variety of ECE classrooms and programs. Given their numerous applications, it is important to establish evidence for measurement invariance of these instruments across different types of ECE classrooms and programs. In this section of the literature review, one empirical study in which measurement invariance was examined for an instrument designed to measure proximal quality indicators in ECE classrooms is described (Downer et al., 2012). A literature search was conducted to identify empirical studies in which measurement invariance was examined for instruments that were designed for use in preschool ECE programs. Of 25 studies that were identified, only one focused on measurement invariance for an instrument designed to measure ECE quality. The majority of studies reviewed focused on instruments designed to assess individual preschool children and were not directly relevant to the present study.

The only study examining measurement invariance of an instrument designed to measure proximal quality indicators across different ECE programs was conducted with the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008). The purpose of this study was to examine measurement invariance for the CLASS across ECE classrooms with different compositions of Latino children and children who were dual language learners (Downer et al., 2012). The analytical sample for the study was 721 preschool classrooms in 11 states that participated in two large-scale studies. Researchers calculated the percentage of dual language learners in each classroom and used these percentages to establish whether classrooms had no dual language

learners, low percentages of dual language learners (i.e., greater than 0%, less than 50%), or high percentages of dual language learners. The same procedures were used to determine whether classrooms had no Latino children, low percentages of Latino children, or high percentages of Latino children. Multiple group confirmatory factor analyses (MG-CFA) were conducted to determine whether the internal structure of the CLASS was invariant with respect to classroom composition. Four models were tested to examine configural, weak, strong, and strict invariance. Findings from this study suggested evidence of strong invariance with respect to classroom composition based on the percentage of dual language learners. The same findings were found in classrooms with different compositions of Latino children. Taken together, these findings provide evidence that the CLASS measures the same constructs in classrooms with different compositions of dual language learners and Latino children and that these constructs are measured on equivalent scales across groups.

The present study was similar to the study conducted by Downer et al. (2012) in four ways. First, it was one of the first empirical studies in which measurement invariance was examined for instruments designed to assess proximal quality indicators. Second, the present study was designed to establish measurement invariance of two widely used measures of the quality in ECE across classrooms with different compositions of children with diverse needs. Third, the methods used to examine measurement invariance of the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989) in the present study were the same as those used by Downer et al. Finally, the analytical sample for the present study was drawn from a large-scale study. The present study and the study described above acknowledge the importance of

establishing measurement invariance of instruments designed to examine proximal quality indicators and extending the literature regarding issues related to measuring quality in ECE.

Early Childhood Environment Rating Scale-Revised

The purpose of this section of the literature review is to provide information about the ECERS-R (Harms et al., 1998). The ECERS-R is a 43-item instrument designed for use in a variety of group programs providing ECE to young children between the ages of 2 and 5 years old (Harms et al., 1998). Items on the ECERS-R are scored on a scale from 1 (*inadequate*) to 7 (*excellent*). Each item is scored based on the presence or absence of quality indicators related to the arrangement of the physical environment, provisions of materials or activities, supervision of children, and interactions between individuals in the classroom. Indicators are anchored at four scale ratings: 1 (*inadequate*), 3 (*minimal*), 5 (*good*), and 7 (*excellent*). A rating of 1 for an item generally indicates that there is no provision in either the physical environment or available materials for the area of quality measured by the item. A rating of 3 indicates the provision of minimal materials, space, or supervision. A rating of 5 indicates adequate materials and space set aside for activities as well as adequate supervision of children. A rating of 7 indicates evidence of teacher planning for individual child differences and encouragement of children's independence within activities in addition to adequate space, materials, and supervision (Harms & Clifford, 1982). Midpoint ratings of 2, 4, and 6 are scored when all of the indicators anchored at the lower anchor are met, but only some of the indicators at the higher anchor are met. Seven subscale scores can be obtained by taking the mean of the item scores for the subscale. An overall score can be obtained by taking the mean of all item scores. Item content, subscale structure,

and scoring guidance of the ECERS-R has evolved to reflect guidelines for developmentally appropriate practice (Bredekamp & Copple, 1997). The following sections of the literature review provide information about (a) the development and revision of the ECERS (Harms & Clifford, 1980), which was the first iteration of the ECERS-R, (b) a review of empirical studies examining the internal structure of the ECERS and ECERS-R, and (c) a review of empirical studies in which the ECERS or ECERS-R was used to examine quality in center-based preschool classrooms where ECE was provided to children with special needs.

Development and Revision of the ECERS

The ECERS (Harms & Clifford, 1980) was originally developed as a self-assessment tool to help ECE program staff examine strengths and weaknesses related to multiple dimensions of quality in ECE. It was designed to capture features of quality of preschool programs through a number of observable indicators assessing the presence or absence of structural and process aspects of quality related to space and furnishings, materials and activities, daily schedule, interactions, and supports for staff (Harms & Clifford, 1980). The ECERS consisted of 37 items organized into seven subscales: (a) personal care routines, (b) Furnishings and Display, (c) Language-Reasoning Experiences, (d) Fine and Gross Motor Activities, (e) Creative Activities, (f) Social Development, and (g) Adult Needs.

A panel of seven experts in the field of ECE reviewed proposed items for the ECERS and rated each item according to its importance in ECE programs. The measure developers revised item content based on expert ratings (Harms & Clifford, 1982). After establishing item content, the ECERS was field-tested by two sets of trained observers in 18 classrooms. Each set of observers included one child

development professional and one person with little or no child development background. Trainers who were unfamiliar with the ECERS but who had been working with teachers in the classrooms in which the measure was field-tested were asked to use a 7-point rating scale similar to that used for the ECERS to rate each classroom on six dimensions that corresponded to the seven ECERS subscales. Rank order correlations for the total ECERS scores across the three groups were relatively high (*range = .70 - .74*; Harms & Clifford, 1982). In addition to efforts to establish content validity of the ECERS (Harms & Clifford, 1980), the measure developers examined inter-rater reliability of total scale scores in 25 classrooms and test-retest reliability of total scale scores in 31 classrooms. Interrater and test-retest reliability were high, with correlations of .88 and .96, respectively (Harms & Clifford, 1982).

The ECERS (Harms & Clifford, 1980) was revised in 1998 (ECERS-R; Harms et al.) to reflect changes in recommendations for developmentally appropriate practice (Bredekamp & Copple, 1997), advancement in the measurement of quality in ECE. Revisions to the ECERS were based on information from three sources: (a) a content analysis of the ECERS and its relationship to other assessments of ECE program quality and documents examining programmatic issues related to quality in ECE, (b) data from research studies using the ECERS, and (c) feedback from ECERS users.

Three focus groups were conducted with practitioners and researchers who had used the ECERS. The purpose of these focus groups was to gather information about how the ECERS functioned in inclusive and culturally diverse settings (Harms et al., 1998). Revisions to the ECERS resulted in the elimination of redundant items, separation of certain items into multiple items, the addition of observable indicators to

some items, the addition of items to address areas not included in the first iteration of the ECERS, and changes to the subscale names and structure (Sakai et al., 2003). This revision process resulted in a revised, 43-item scale that is organized into seven subscales: (a) Space and Furnishings, (b) Personal Care Routines, (c) Language-Reasoning, (d) Activities, (e) Interaction, (f) Program Structure, and (g) Parents and Staff.

The ECERS-R was field-tested in 21 classrooms to assess interrater score reliability and to examine the internal consistency of items. Interrater score reliability was comparable to that of the original measure, with a Pearson product correlation of .92 and Spearman rank order correlation of .86 for total scale scores (Harms et al., 1998). Internal consistencies for subscales ranged from .71 - .88, with a total scale internal consistency score reliability of .92 (Harms et al., 1998). Initial field tests of the ECERS and ECERS-R provided some promising evidence about the content validity and reliability of scores obtained from these instruments, but did not provide any information about the internal structure of these instruments.

Although it was originally developed for use as a self-assessment tool, the ECERS (Harms & Clifford, 1980) and the ECERS-R (Harms et al., 1998) have become some of the most widely used instruments to evaluate quality in ECE classrooms and programs. Before it was revised, the ECERS was used widely in research studies to evaluate the quality in ECE across several types of ECE programs and classrooms. In addition, the ECERS-R is used widely to monitor quality in ECE programs, as the basis to provide incentives to high-quality ECE programs, and to communicate quality ratings to consumers of ECE, and to guide program development endeavors (ERSI, n. d.).

Despite the fact that the consequences for some of these uses involve very high stakes for ECE programs, there have been very few studies to validate the ECERS-R for such purposes.

Structural Validity Evidence for the ECERS/ECERS-R

Although the developers of the ECERS (Harms & Clifford, 1980) and the ECERS-R (Harms et al., 1998) asserted that the scales measure multiple dimensions of quality in ECE classrooms and programs, there has been debate about the extent to which the items of these scales discriminate among different features of quality. Principal component and factor analyses of ECERS and ECERS-R data have yielded inconsistent findings, but there have been no empirical studies conducted that have corroborated the internal structure proposed by the instrument developers.

The purpose of this section of the literature review is to describe the approaches used previously to examine the factor structure of the ECERS and ECERS-R and summarize findings from these studies. Articles, studies, and reports included in the literature review for this section met the following criteria: (a) principal components analysis (PCA), exploratory factor analysis (EFA), confirmatory factor analysis (CFA), or multiple group factor analysis (MG-FA) was conducted; (b) the measure of interest was the English version of either the ECERS (Harms & Clifford, 1980) or the ECERS-R; (c) the authors provided information about the methodology used to conduct the analyses; (d) the authors' description of the internal structure of the scale included, at a minimum, an explanation of pattern of item loadings onto identified factors. The latter two criteria were applied because a number of studies briefly mentioned factors identified from a principal components or factor analysis, but they did not provide sufficient information to conduct a critical analysis of the methodology or findings from such examinations. Of

ten articles that were located, a total of seven articles and reports met the criteria for inclusion in the present literature review. Tables 2-2 and 2-3 display information about the methodologies and study findings from empirical studies in which the factor structure of the ECERS or the ECERS-R was examined. The studies reviewed in this portion of the literature review are grouped according to whether the measure of interest was the ECERS or the ECERS-R. A brief description of each study is provided, followed by a summary of findings across studies.

Empirical studies examining the factor structure of the ECERS

Three studies used principal components or factor analyses to examine the factor structure of the ECERS (Harms & Clifford, 1980). Findings from two of the three studies reviewed suggested that the ECERS measures two dimensions of ECE. Whitebook et al. (1989) were the first researchers to publish results from a factor analysis with the ECERS. In this study, data from 313 classrooms that participated in the NCCSS were used. Using a maximum likelihood extraction with oblique rotation to identify factors, these authors proposed two dimensions of quality measured by the ECERS. The first dimension was labeled Appropriate Caregiving and was measured by 16 items. These items included all items that were designed to capture teacher-child interactions, supervision of children, and discipline (i.e., understanding language, using language, reasoning, informal language, supervision of fine, gross motor, and creative activities, tone of interactions) as well as items related to personal care routines (i.e., greetings/departure, meals/snacks, nap/rest, diapering/toileting) and scheduled activities (i.e., schedule of creative activities, music/movement activities, free play, group time). Factor loadings for the Appropriate Caregiving factor ranged from .57 (diapering/toileting) to .83 (using language), with a majority of factor loadings above .70.

A second factor, Developmentally Appropriate Activity, consisted of 10 items that pertained to the physical environment, including arrangement of physical space and provision of materials and activities. Factor loadings ranged from .51 (cultural awareness activities) to .85 (room arrangement). Using the definitions provided by Cassidy, Hestenes, Hansen, et al. (2005), the Appropriate Caregiving factor appears to measure primarily process quality, whereas the Developmentally Appropriate Activity factor appears to measure primarily structural quality.

Sakai et al. (2003) identified a similar factor structure for the ECERS (Harms & Clifford, 1980), despite using a much smaller sample size ($N = 68$) and different techniques for the extraction and rotation of factors. A principal components analysis with orthogonal rotation was conducted using 33 of the 37 items on the ECERS. Two factors that represented developmentally appropriate caregiving and developmentally appropriate activities accounted for 45% of the total variance in the data. The item composition for these factors was generally similar to the item composition for the Appropriate Caregiving and Developmentally Appropriate Activity factors identified by Whitebook et al. (1989); however, the factor structure identified by Sakai et al. resulted in fewer items per factor and a larger number of items that did not load onto either factor ($v = 7$).

Findings from the two studies previously described were not corroborated by a study conducted by Scarr et al. (1994). Maximum likelihood estimation was used to extract factors using subscale- and item-level ECERS (Harms & Clifford, 1980) data from 120 classrooms. Both orthogonal and oblique rotation yielded a single factor that accounted for 69% of common variance. Given the evidence for one large factor and

the fact that all items loaded on the factor, Scarr et al. created three parallel forms of the ECERS, each with 12 randomly selected items. Despite the range of content addressed in ECERS, composite scores on each of the parallel forms were correlated extremely highly with the total scale score (*range* = .93-.95) and had internal consistency score reliability estimates ranging from .84-.92.

Empirical studies examining the factor structure of the ECERS-R

Five of the studies reviewed for this portion of the literature review examined the factor structure of the ECERS-R (Harms et al., 1998). All of these studies provided some empirical support for at least a two-factor internal structure that differentiates structural quality from process quality (see Table 2-3). Sakai et al. (2003) sought to determine whether the two-factor structure they and others (Whitebook et al., 1989) identified for the ECERS (Harms & Clifford, 1980) could be replicated with the ECERS-R. Using the same sample of 68 classrooms they used to examine the internal structure of the ECERS, these authors conducted a principal components analysis with orthogonal rotation. The item pool for this analysis consisted of all the items from the ECERS-R except those from the Parents and Staff subscale ($v = 6$). A two-factor solution accounted for 67% of total variance in the data. Although there was a general pattern of items related to activities and materials loading on one factor and items related to teacher-child interactions loading on the second factor (see Table 2-4), the results of this analysis were less interpretable than those for the ECERS due to the reversal of items that were differentiated clearly in analyses of ECERS data (e.g., staff-child interactions). The researchers who conducted this study posited that had the sample been larger, the factors might have been better differentiated.

In a study using a significantly larger sample of 240 classrooms, Clifford et al. (2005) found that items from the ECERS-R (Harms et al., 1998) mapped clearly onto two factors similar to those previously described. Clifford et al. reported using data from 36 ECERS-R items to conduct an exploratory factor analysis with orthogonal rotation. The method used to extract the factors (e.g., principal components, maximum likelihood) was not specified. Findings from this study yielded a two-factor solution that included 23 of the 36 items included in the analysis. The item composition of each of these factors is presented in Table 2-4. One factor was labeled Teaching and Interactions and included 11 items designed to capture primarily teacher-child interactions, supervision of children, and interactions between children another factor labeled. The second factor, labeled Provisions for Learning, included 12 items related to physical space and the provision of activities and materials.

Cassidy, Hestenes, Hegde, et al. (2005) also reported two dimensions of quality measured by the ECERS-R (Harms et al., 1998). Data for this study came from 1313 classrooms that participated in a QRIS in North Carolina. Three methods of extraction (i.e., principal components, principal factor, maximum likelihood) were used to conduct EFA with orthogonal rotation with a sample of 486 classrooms from the larger sample. In addition, three CFA models were tested with a sample of 472 classrooms from the larger sample. The models examined were two- and three-factor models informed by the EFAs and a seven-factor model representing the subscales proposed by the developers of the ECERS-R.

Findings from these analyses suggested a three-factor model; however, one factor was dropped due to having a limited number of items ($v=2$) and a low correlation

with the total ECERS-R score (Cassidy, Hestenes, Hegde, et al., 2005). The final two-factor model adopted by the authors had one factor consisting of nine items and measured the quality of activities and materials provided in the classroom. A second factor included seven items and captured the quality of language and interactions occurring in the classroom. The items associated with each factor are presented in Table 2-4. Internal consistency score reliabilities for the two factors were .88 and .81, respectively for the sample of 472 classrooms. When the samples from the exploratory and confirmatory analyses were combined ($N = 958$), internal consistency score reliabilities for the two factors were .87 and .81, respectively.

In a follow-up study, Cassidy, Hestenes, Hansen, et al. (2005) applied their definitions of structural and process quality to the indicators comprising the 16-item scale and determined that 83% of the indicators in the Activities/Materials subscale were indicators of structural quality and 90% of the indicators in the Language/Interactions subscale were indicators of process quality.

Two additional studies (Gordon et al., 2013; Perlman et al. 2004) yielded evidence for a three-factor internal structure of the ECERS-R. Although the number of factors was similar, the constructs and pattern of items associated with each factor differed across studies (see Table 2-4). Perlman et al. (2004) conducted a factor analysis using maximum likelihood extraction and oblique rotation with a sample of 326 classrooms. The first factor explained 71% of the common variance. Fifteen items related to the arrangement of the physical space and provision of materials and activities loaded on this factor. The second factor explained 10% of the common variance and included 10 items related to the staff-child interactions and supervision of

children. The final factor explained 6% of the common variance and included 8 items related to provisions for parents and staff. Despite the statistical evidence for a three-factor structure, Perlman et al. asserted that the ECERS-R only measures one global construct of quality. The rationale for this assertion was that the eigenvalue for the first factor was substantially larger than that of the second two factors. The authors also posited that because the correlations between factors ranged from .50 to .62, all three factors measured similar constructs.

The most recent study published that examined the structural validity ECERS-R (Harms et al., 1998) is particularly relevant to the present study, because the authors used data from the ECLS-B (Gordon et al., 2013). Gordon et al. (2013) conducted EFAs with oblique rotation using data from 1350 classrooms that were observed during the ECLS-B. Six items from the Provisions for Staff and Parents subscale were not included in their analyses, because these data were not collected as part of the ECLS-B observation protocol. In contrast to Perlman et al. (2004), these authors did not find evidence for a one-factor model. Although the authors reported that a six-factor model was feasible based on model fit statistics, the item-loadings for the six factors were not consistent with those proposed *a priori* by the developers of the ECERS-R. Furthermore, there were multiple factors on which only two or three items loaded. Although it included an additional factor compared to previous studies, the three-factor solution was most consistent with reports from previous researchers and was also defensible theoretically. The first factor was composed of 19 items related to the provision of space, materials, and activities. Seven items related to personal care routines (e.g., health practices, safety practices, diapering/toileting) had salient loadings

for the second factor. The third factor was composed of nine items related to language-reasoning and interactions. As shown in Table 2-3, the first and third factors derived from this study are very similar to those reported in other studies examining the structural validity of the ECERS-R.

Summary of empirical studies examining factor structure of the ECERS/ECERS-R

Although the item composition of the factor structures proposed by researchers who have examined the internal structure of the ECERS (Harms & Clifford, 1980) and the ECERS-R (Harms et al., 1998) differ slightly, a consistent finding is that these instruments measure at least two dimensions of quality (Cassidy, Hestenes, Hegde, et al., 2005; Clifford et al., 2005; Gordon et al., 2013; Sakai et al., 2003; Whitebook et al., 1989). When the definitions proposed by Cassidy, Hestenes, Hansen, et al. (2005) are applied to the items composing the empirically derived subscales from these studies, there appears to be a consistent finding that one subscale measures primarily structural quality and the other subscale measures primarily process quality. No studies have been conducted that have corroborated assertions of the measure developers that these instruments measure seven dimensions of quality in ECE classrooms. For many of the studies reviewed (Cassidy, Hestenes, Hegde, et al., 2005; Clifford et al., 2005; Perlman et al., 2004; Scarr et al., 1994), the proposed factor structure resulted in a loss of several items that were originally included in the model (see Table 2-4). These findings provide support for the use of a shortened form to examine the dimensions of structural and process quality in ECE programs. It is important to note, however, the items commonly excluded from these shortened forms are related to health and safety practices or personal care routines, which are often among the items that receive the lowest scores. The most recent study conducted to examine the internal structure of

the ECERS-R (Gordon et al., 2013) yielded a factor structure that (a) retained all of the items that were originally included in the analysis, (b) corroborated evidence from previous studies that suggested the ECERS and ECERS-R measure both structural and process quality in ECE programs, and (c) included a factor composed of items related to personal care routines.

Several limitations of the studies reviewed in this portion of the literature review should be noted. First, the sample sizes for some of the studies reviewed ranged from 68 classrooms to 1350 classrooms. Some of these sample sizes were relatively small in relation to the number of observed variables that were included in the statistical analyses. The variation in sample size might have contributed to differing patterns of item loadings across studies. Second, many of the authors reported using an orthogonal rotation (Cassidy, Hestenes, Hegde, et al., 2005; Clifford et al., 2005; Sakai et al., 2003; Scarr et al., 1994), which assumes that quality dimensions are uncorrelated. This methodology is inconsistent with conceptualizations of quality in ECE that stress the relationships between different features of quality (Cryer, 1999; Love et al., 1996; Vandell & Wolfe, 2000). Third, for most of the studies reviewed, the criteria for identifying the most favorable factor structure were based on eigenvalues, proportion of variance explained, scree plots, and interpretability of results. Although the application of both statistical and theoretical criteria is aligned with current recommendations for factor analytic techniques, additional statistical criteria were available that would have provided more information about the fit of the models to the data that were analyzed. For example, the assertion that the ECERS and ECERS-R measure only one dimension of quality (Scarr et al., 1994; Perlman et al., 2004) was

based primarily on the examination of eigenvalues and correlations between subscale and factor scores rather than indices of model fit. When one-, three-, and six-model factors were examined using multiple indices of model fit, there was evidence to suggest that both the three- and six-model solutions produced a better fit to the data than the one-factor model (Gordon et al., 2013). When these statistical criteria were combined with theoretical criteria, the three-factor model was shown to be feasible both statistically and theoretically.

A final limitation of existing research examining the structural validity of the ECERS (Harms & Clifford, 1980) and ECERS-R (Harms et al., 1998) is the lack of empirical evidence to support the assertion that these instruments measure the same dimensions of quality in the same manner across qualitatively different center-based preschool ECE settings. Although the developers of the ECERS asserted that the instrument was designed to capture features of quality that are present in all ECE programs (Harms & Clifford, 1982), no studies have been designed to examine the accuracy of this assertion. Given the diversity of ECE programs in which these instruments have been used, it is important to conduct empirical studies to test this assumption. Without evidence that these instruments measure the same constructs in the same manner across the different situations in which they have been used, it is not possible to know whether the scores are comparable across situations.

Use of ECERS/ECERS-R to Examine Quality in Classrooms Enrolling Children with Special Needs

The purpose of this section of the literature review is to synthesize findings from empirical studies using either the ECERS (Harms & Clifford, 1980) or the ECERS-R (Harms et al., 1998) to characterize quality in center-based preschool classrooms where

children with special needs were enrolled. Studies reviewed in this portion of the literature review met the following criteria: (a) data were collected in preschool classrooms; (b) the sample included children with special needs or classrooms that provided ECE services to these children; (c) either the ECERS or ECERS-R was used to characterize quality in ECE classrooms or programs. A total of nine studies published in eight articles were located and selected for review. The studies reviewed have direct relevance to the present study because they provide information about how the ECERS and ECERS-R have been used to characterize quality in ECE programs that provide services to children with special needs. Findings from these studies are important because they represent part of the evidence base used to inform research, policy, and practice in ECE and early childhood special education.

Each of the studies reviewed examined comparisons in the quality in ECE across ECE classrooms or programs with differing compositions of children with special needs. Table 2-5 provides a description of the terminology and criteria used to define groups of ECE classrooms. As shown in Table 2-5, a common comparison for the studies reviewed was of inclusive and noninclusive classrooms. Although the criteria used to define inclusive and noninclusive classrooms differed slightly across the studies presented in Table 2-5, these two types of ECE classrooms are similar, respectively, to ECE classrooms characterized as INC and NSN in the present study. Given the similarities in criteria used to define these two types of classrooms, the terms INC classrooms and NSN classrooms are used in the present review to describe classrooms previously described as inclusive or noninclusive by study authors.

In addition to summarizing the terminology and criteria used to define different types of ECE classrooms, Table 2-5 shows the unit of analysis and level of inference (e.g., program, classroom, or individual child) for each study. The unit of analysis for every study reviewed was ECE classrooms. Although several researchers used scores based on observations of ECE classrooms to make inferences about programs or individual children, findings from these studies are discussed with respect to the unit of analysis (i.e., ECE classrooms).

In this portion of the literature review, findings from the studies noted in Table 2-5 are reviewed and synthesized. A brief description of each study is provided, followed by a summary of findings across studies and a discussion of limitations and gaps in existing research examining the quality of classrooms where ECE is provided to preschool children with special needs. Studies are grouped according to whether the ECERS (Harms & Clifford, 1980) or the ECERS-R (Harms et al., 1998) was used to characterize the quality in ECE.

Use of ECERS to examine the quality in ECE for children with special needs

There were three studies in which the ECERS (Harms & Clifford, 1980) was used to characterize the quality in ECE for preschool children with special needs (Bailey et al., 1982; Buysse et al., 1999; La Paro et al., 1998). All three studies used the ECERS as a dependent measure to make comparisons of quality across two types of ECE programs; however, the two comparison groups differed across studies. One study compared the quality in ECE in preschool classrooms in which only children with special needs were enrolled to ECE classrooms in which no children with special needs were enrolled (Bailey et al., 1982). One study (La Paro et al., 1998) made comparisons across center-based INC classrooms and center-based preschool ECE classrooms in

which only children with special needs were enrolled. The third study involved comparison of quality across INC and NSN classrooms. Studies are reviewed in the order in which they were published to highlight an important shift in research foci across the time span in which these studies were conducted.

The first study (Bailey et al., 1982) focused on differences in the quality in ECE in classrooms in which primarily to children with special needs were enrolled and classrooms in which primarily children without special needs were enrolled. Total sum scores for the ECERS in these classrooms were compared. Findings from this study suggested global quality was higher in classrooms in which no children with special needs were enrolled than in classrooms where primarily children with special needs were enrolled [$t(-78) = 3.46, p = .001$]. A follow-up discriminant analysis revealed six items on the ECERS collectively made the highest contribution to the observed differences: (a) blocks, (b) provisions for exceptional children, (c) art, (d) space to be alone, (e) scheduled time for gross motor, and (f) cultural awareness.

Several years later, La Paro et al. (1998) examined differences in the global quality of 29 ECE classrooms in which only children with disabilities were enrolled and 29 INC classrooms. Data from a number of instruments were used to characterize quality in these settings. In addition to the ECERS, these authors administered the Classroom Practices Inventory (CPI; Hyson et al., 1991), which is an observational instrument designed to measure the extent to which preschool teachers implement developmentally appropriate practices. Data from self-report instruments were used to characterize teacher beliefs about developmentally appropriate practices and instructional supports and implementation of developmentally appropriate practices.

Scores from these instruments were used in a descriptive discriminant analysis to determine which combination of quality indicators accounted for the largest proportion of variance in the quality in ECE between the two types of classrooms. Findings from this study revealed that only 14.6% of variance in the linear composite of quality indicators was accounted for by classroom type. Furthermore, there were no significant or notable differences in overall mean scores on the ECERS.

Buysse et al. (1999) conducted a similar study comparing global quality in 62 INC classrooms and 118 NSN classrooms. Classrooms were sampled randomly from ECE programs. Observers rated all of the ECERS items except those related to adult needs and compared the total mean scores across INC and NSN classrooms. Results from an ANCOVA revealed INC classrooms had higher mean scores on the ECERS than NSN classrooms ($p < .01$). Although these findings appeared to provide promising evidence that children with disabilities receive access to high-quality ECE experiences, the authors noted the extent to which these findings were meaningful was confounded by the fact that there was no evidence to suggest that the ECERS was an appropriate instrument with which to assess quality in inclusive programs.

The shift in research focus represented by three studies described reflects changes in policy and recommended practices for ECE that occurred over the time span in which the studies were conducted. These changes highlighted the importance of providing ECE services for preschool children with disabilities in inclusive settings. When Bailey et al. (1982) conducted their study, federal mandates for the provision of special education and related services to preschool children with disabilities in LRE appropriate had not yet been codified into law. Nearly 20 years later, these mandates

had been in place for a number of years. In addition, guidelines for developmentally appropriate practice in ECE programs had been revised to reflect the need to provide ECE services to children with diverse needs (Bredekamp & Copple, 1997). The studies conducted by La Paro et al. (1998) and Buysse et al. (1999) were the first of several studies that would examine the quality in ECE in INC classrooms. The changing landscape for the provision of ECE services also prompted the revision of the ECERS at the same time these two studies were being conducted. During the revision process, there was a particular emphasis on ensuring that the ECERS-R was useful for capturing quality in inclusive preschool programs.

Use of ECERS-R to examine quality in ECE classrooms for children with special needs

Six studies have been conducted in the past 10 years that have used the ECERS-R (Harms et al., 1998, 2005) as a dependent measure to make comparisons of quality in INC and NSN classrooms (Clawson & Luze, 2008; Grisham-Brown et al., 2010; Hestenes et al., 2008; Knoche et al., 2006; Wall et al., 2006). Although the primary research questions were similar across studies, data collection and sampling procedures varied. Studies are reviewed such that findings from studies using similar sampling and data collection procedures are presented together.

Four studies used similar sampling and data collection procedures as those used in the studies described previously (Grisham-Brown et al., 2010; Hestenes et al., 2008; Knoche et al., 2006). For these studies, classrooms were sampled from programs and data were not linked to children in the classrooms. Knoche et al. (2006) examined differences in ECERS-R (Harms et al., 1998) subscales and overall mean scores for 58 INC and 54 NSN classrooms. All subscale scores and the overall mean score for the

ECERS-R were higher on average in INC classrooms, but none of these differences were statistically significant. P values were $p < .10$ for the Space and Furnishings subscale ($t = 1.81$, $d = 0.37$), the Language and Reasoning subscale ($t = 1.83$, $d = 0.36$), the Interaction subscale ($t = 1.67$, $d = 0.33$), and the overall mean score ($t = 1.93$, $d = 0.37$).

In a similar study with a comparable sample size, Grisham-Brown et al. (2010) reported statistically significant results that corroborated findings from Knoche et al. (2006). Secondary analyses of 33 INC and 33 NSN classrooms selected to participate in a statewide evaluation via stratified, proportionate random sampling were conducted. Grisham-Brown et al. reported higher subscale and overall mean scores on the ECERS-R for inclusive classrooms compared to non-inclusive classrooms. Findings were statistically significant for Space and Furnishings ($t = 2.98$, $p < .01$, $d = .74$), Language and Reasoning ($t = 2.03$, $p < .05$, $d = 0.50$), Activities ($t = 3.84$, $p < .001$, $d = 0.96$), Program Structure ($t = 6.40$, $p < .001$, $d = 1.60$), and overall mean score ($t = 3.54$, $p < .01$, $d = 0.88$).

Hestenes et al. (2008) conducted two studies examining differences in the quality in ECE in INC and NSN classrooms. The first study included data from a large sample of classrooms participating in a statewide QRIS in North Carolina. In addition to reporting differences on ECERS-R subscale and overall mean scores for the 459 INC classrooms and 854 NSN classrooms observed during this study, the authors reported differences on two factor scores (i.e., Language/Interaction, Activities/Materials). Factor scores were identified via EFA and were comprised of 16 of the original 43 items from the ECERS-R (Cassidy, Hestenes, Hegde, et al., 2005). Scores for both factors, the

ECERS-R subscales, and the mean overall score on the ECERS-R were higher on average in INC classrooms. Findings were statistically significant for the Activities/Materials factor [$F(1,1311) = 18.9, p < .001$], the Language/Interactions factor [$F(1, 1311) = 49.5, p < .001$], and for total mean score [$F(1,1311) = 42.4, p < .001$]. Mean differences on subscale scores were statistically significant ($p < .001$) for all subscales except Personal Care Routines and Space and Furnishings.

The second study conducted by Hestenes et al. (2008) was conducted with a significantly smaller sample of 20 INC classrooms and 24 NSN classrooms. Although this study included comparisons of ECERS-R factor scores, ECERS-R subscale scores, and ECERS-R total score, the primary focus of the study was on process quality. An observational rating scale was used to measure the frequency, quality, and appropriateness of teacher-child interactions. No statistically significant differences were found for the ECERS-R factor scores, ECERS-R subscale scores, or overall mean score on the ECERS-R, but there were statistically significant findings to suggest the quantity and quality of teacher-child interactions was higher in INC classrooms. The lack of statistically significant findings for the ECERS-R might be due, in part, to the small sample size of this study. Given other observational rating scale used was designed to measure teacher-child interactions with more specificity than the ECERS-R, it is possible this instrument was more sensitive to detect differences in smaller sample sizes than the ECERS-R.

In contrast to the studies previously described, two of the studies reviewed used scores from the ECERS-R as a proxy to represent the quality in ECE experienced by individual children enrolled in the classroom. The first study (Wall et al., 2006) included

data from 91 children with special needs and 337 children without special needs. The sample for this study was a subsample from the EHSRE project, which was an experimental study conducted to examine the impact of Early Head Start on children and families. For this study, data from children and families in the control and experimental groups were combined, and comparisons of the quality in ECE for children with and without special needs were made. Children were characterized as having special needs if they met at least one of the following criteria: (a) received early intervention or preschool special education services, (b) had diagnosed conditions that would make them eligible for early intervention or preschool special education services or that were likely to be associated with developmental delays (e.g., sensory impairment, chromosomal abnormality), (c) had suspected delays (e.g., difficulty using arms or legs), as reported by parents or professionals, or (d) had biological risks (e.g., low birth weight). A comparison of overall mean scores on the ECERS-R did not yield statistically significant differences between groups; however, mean scores were higher on average for children with special needs.

A more recent study (Clawson & Luze, 2008) focused on the quality of individual ECE experiences of 30 children with special needs and 30 children without special needs. For this study, items from the ECERS-R were adapted and sum scores for three scales proposed by the authors were used to characterize the quality of interaction experiences, language experiences, and curriculum experiences for individual children who participated in the study. The rationale for adapting ECERS-R items to be used for individual children was that classroom-level ratings provided information about the quality in ECE an individual child might potentially experience, but not information about

what was actually experienced. Findings did not reveal statistically significant differences in item, subscale, or total scores on the ECERS-R, nor did they detect statistically significant differences on the individualized ECERS-R subscales.

Summary of research examining the quality in ECE for children with special needs

Several trends emerged from a review of empirical studies in which the ECERS or ECERS-R were used to examine quality in center-based preschool ECE programs or classrooms in which children with special needs were enrolled. First, there has been a clear shift in research focus from examining global quality in center-based preschool ECE programs or classrooms where ECE is provided to either children with special needs or children without special needs to examining the quality in ECE in INC classrooms. Second, there appears to be an emerging interest in using the ECERS-R as a measure of quality for individual children. In general, findings from empirical studies comparing the quality of INC and NSN classrooms suggest global quality is higher in INC classrooms. Findings related to differences in subscale scores varied across studies, with relatively consistent findings that scores for the activities, language and reasoning, and interactions subscales were higher in inclusive classrooms. When ECERS-R scores were used as proxies for the quality in ECE experienced by individual children, no differences were found in the quality of experiences for children with and without special needs (Clawson & Fuze, 2008; Wall et al., 2006).

Several limitations of the studies reviewed are noted. First, there is no empirical evidence that scores from ECERS (Harms & Clifford, 1980) or the ECERS-R (Harms et al., 1998) are equivalent in the different ECE classrooms in which they were administered for these studies. Researchers who used the ECERS in their

investigations expressed concerns about the validity of scores from this instrument in classrooms where ECE services were provided to children with special needs (Bailey et al., 1982; Buysse et al., 1999). Some researchers suggested that the ECERS-R was the only measure of quality that had been validated for use in inclusive ECE programs and classrooms (Hestenes et al., 2008). Although it is true that the processes used to establish content validity of the ECERS-R included procedures to ensure that the quality indicators on the ECERS-R were applicable to inclusive classrooms (Harms et al., 1998), there were no published empirical studies identified during the present review of literature providing information about whether these instruments measure similar constructs in INC and NSN classrooms or whether the scores obtained from these instruments are equivalent across these contexts. Without such evidence, it is impossible to interpret the findings of the studies reviewed in this section of the literature review with any degree of certainty.

A second limitation applies to several of the studies in which the ECERS-R (Harms et al., 1998, 2005) was used and is related to the use of scores on subscale that have not been empirically validated as dependent variables to examine differences in ECE quality between INC and NSN classrooms. Although findings from studies examining the factor structure of the ECERS-R were mixed with regard to how many dimensions of quality are measured by this instrument, none of the nine studies reviewed provided evidence for the seven subscales originally proposed by the instrument developers. More appropriate comparisons would have been comparisons of empirically validated factor scores. Even assuming that there was measurement invariance of the ECERS-R across the contexts in which they were administered,

reports of differences on subscale scores should be interpreted with caution, given the lack of evidence regarding measurement invariance for the instrument.

Additional limitations regarding the research design of the studies reviewed might have accounted for some of the inconsistencies in findings across studies. Many of the studies reviewed had relatively small sample sizes (Bailey et al., 1982; Clawson & Luze, 2008; Grisham-Brown et al., 2010; Knoche et al., 2006; La Paro et al., 1998;) or uneven cell sizes (Buysse et al., 1999; Hestenes et al., 2006; Wall et al., 2006), which resulted in limited power to detect statistically significant differences between comparison groups. When quality of INC classrooms was compared to that of NSN classrooms, researchers generally reported higher quality in INC classrooms, with varying levels of statistical significance. For studies with small sample sizes or uneven cell sizes, it is possible that the loss of power due to sample size or uneven cell sizes accounted for the lack of statistically significant findings. In some cases, the ability to detect differences in quality across comparison groups was hindered by limited variability in quality scores. Several of the studies reported restricted ranges in quality, or systematically higher quality scores compared to those reported by other researchers. In these cases, the absence of detectable variability in quality across comparison groups might actually be due to a lack of variability in overall quality.

The present study was designed to address the noted limitations in a number of ways. First, the purpose of the present study is to establish empirical evidence regarding measurement invariance of the ECERS-R (Harms et al., 1998). Findings from the present study provide information about the interpretability of findings from previous studies in which this instrument was used to characterize the quality in ECE in

INC classrooms. Second, the confirmatory model used to examine measurement invariance of the ECERS-R for the present study included three empirically validated factors. Finally, the analytic sample for the present study consisted of a large sample of center-based preschool ECE classrooms in which a nationally representative sample of children who received center-based ECE services for more than 10 hours per week were enrolled. The analytical sample for the present study allowed for comparisons in the quality in ECE across INC and NSN classrooms that were not limited by small sample sizes or restricted variability in The quality in ECE.

Arnett Caregiver Interaction Scale

In the this section of the literature review, information about the CIS (Arnett, 1989) is provided. Information about the development of the CIS is provided, followed by a review of empirical studies designed to establish structural validity evidence for the CIS. A review of empirical studies in which the CIS was used to characterize quality in ECE programs or classrooms where children with special needs were enrolled concludes this section of the literature review.

Development of the CIS

The CIS is a judgment-based observational rating scale designed to provide a rating of caregiver interactions in day care centers (Arnett, 1986; Arnett, 1989). It is composed of 26 items representing caregiver behaviors. Items are scored on a scale from 1 (*not at all*) to 4 (*very much*). The rating scale for the CIS signifies the extent to which the caregiver engaged in the behavior described on the CIS. As shown in Table A-2, items on the CIS are either positively (e.g., speaks warmly to children) or negatively (e.g., seems critical of the children) oriented and are hypothesized to measure four dimensions of caregiver interactions: (a) sensitivity, (b) harshness, (c)

detachment, and (d) permissiveness. Total and subscale composite scores represent the sum or average of item-level scores, with items that are negatively oriented reverse-scored. The CIS was developed during pilot observations in Head Start centers and was field-tested until three observers achieved 80% interobserver agreement for all items (Arnett, 1986; 1989). The subscales proposed by the author were based on principal component analyses with data from 59 classrooms. The CIS has been used widely in large-scale research studies (e.g., Helburn, 1995; IES, NCES, n.d.; Layzer et al., 1993; OPRE, 1996-2010; OPRE, 1997-2013; Whitebook et al., 1989) to characterize the quality of caregiver interactions in ECE classrooms; however, there have been very few studies in which the psychometric properties of the instrument have been examined. One purpose of the present study was to establish validity evidence for the CIS based on internal structure. Specifically, measurement invariance of the instrument was examined across inclusive and non-inclusive center-based preschool classrooms. To date, there have been no empirical studies published that have examined measurement invariance of the CIS, and very few studies that have examined the internal structure of the instrument.

Empirical Studies Examining Factor Structure of the CIS

In this portion of the literature review, the approaches used previously to examine the factor structure of the CIS are described. Articles, studies, and reports included in the following review met the following criteria: (a) principal components analysis (PCA), exploratory factor analysis (EFA), confirmatory factor analysis (CFA), or multiple group factor analysis (MG-FA) was conducted; (b) the measure of interest was the CIS (Arnett, 1989); (c) the authors provided information about the methodology used to conduct the analyses; (d) the authors' description of the internal structure of the scale

included, at a minimum, an explanation of pattern of item loadings onto identified factors.

Only two studies were identified that met the criteria described above. Two additional technical reports were located that mentioned conducting analyses to examine the factor structure of the CIS (Arnett, 1989), but there was not sufficient information for a critical analysis of these studies.

Arnett (1986, 1989) conducted the first study to examine the factor structure of the CIS. Findings from this study were the basis for the four subscale scores that are commonly reported for CIS. A principal components analysis with oblique rotation was conducted with data from 59 classrooms in Bermuda. Examination of eigenvalues and image analyses resulted in a four-factor structure that accounted for 66% of total variance in item-level scores. The subscales originally proposed were (a) Positive Interaction ($v = 10$), (b) Punitiveness ($v = 9$), (c) Detachment ($v = 4$), and (h) Permissiveness ($v = 4$). Table A-1 displays the items that were associated with each of these subscales. When the CIS was administered in subsequent studies, two of the subscales (i.e., Positive Interaction, Punitiveness) were renamed, but the proposed internal structure remained the same. A number of technical reports for large-scale studies in which the CIS was administered have reported a similar factor structure for the instrument (Layzer et al., 1993; Love, Ryer, & Faddis, 1992; Whitebook et al., 1989).

The most recent published study in which the internal structure of the CIS (Arnett, 1989) was examined did not corroborate reports from previous studies. Colwell et al. (2012) used CFA, EFA, and item response theory models to examine the internal structure of the CIS with data from 1350 center-based preschool classrooms that were

observed in the ECLS-B preschool data collection wave. In contrast to the analytical techniques employed by Arnett (1989), these methodologies provide information about the interrelationships between items to represent latent variables. Colwell et al. (2012) proposed that item correlations reported in previous studies might have been confounded by methodological constructs rather than substantive constructs. Specifically, these authors posited that the negative or positive orientation of the item wording might contribute to item correlations. To account for this possibility, bi-factor latent models were tested to separate methodological factors from substantive factors. When methodological factors were taken into account, there was not evidence to suggest that the CIS is a multidimensional measure of caregiver interactions.

The work of Colwell et al. (2012) was used as a basis for the MG-CFA of CIS (Arnett, 1989) data in the present study. Although there have been previous reports providing validity evidence that the CIS is a multidimensional measure of caregiver interactions, these studies have several limitations that should be noted. First, only one report provided enough information about the analyses or findings to support a critical analysis of the study. Arnett (1989) conducted a principal components analysis, which is a data reduction technique that does not provide information about interrelationships between items to represent latent variables. In addition, this study was conducted with a limited sample of classrooms outside the United States.

In contrast, Colwell et al. used multiple latent variable modeling techniques with large sample of classrooms in the United States. Furthermore, the analytic sample for the study conducted by Colwell et al. was the same as the analytic sample for the present study.

Findings from the present study extend previous studies in which the factor structure of the CIS was examined. The present study provides information about the extent to which the CIS measures the construct of caregiver interactions equally in INC and NSN classrooms. Findings from the present study are important because the CIS has been used to characterize the quality of caregiver interactions in these settings for a number of large-scale studies. In addition, CIS data from such studies have been used to make comparisons of the quality of caregiver interactions in across these settings (Knoche et al., 2006; Wall et al., 2006). Findings from the present study provide information about the extent to which it is appropriate to use the CIS to make inferences about the quality of caregiver interactions in INC and NSN classrooms.

Use of the CIS to Characterize Process Quality in Inclusive Preschool Classrooms

The purpose of this section of the literature review is to synthesize findings from empirical studies in which the CIS (Arnett, 1989) was used to characterize quality in classrooms where children with special needs were enrolled. Studies reviewed in this portion of the literature review met the following criteria: (a) data were collected in center-based preschool classrooms; (b) the sample included children with special needs or classrooms that provided ECE services to children with special needs; (c) either the CIS was used to characterize the process quality. A total of two studies were identified. The studies reviewed have direct relevance to the present study because they provide information about how the CIS has been used to characterize quality in center-based preschool ECE programs or classrooms. Findings from these studies are important because they represent part of the evidence base used to inform research, policy, and practice in ECE and early childhood special education.

Both of the studies identified for review also used the ECERS-R (Harms et al., 1998) to characterize the quality in ECE provided in INC classrooms and have been described previously in the present literature review. For the first study, classrooms were sampled from programs, and data were used to characterize global, structural, and process quality (Knoche et al., 2006). Knoche et al. (2006) examined differences in CIS overall mean scores for 143 INC classrooms and 206 NSN classrooms. The sample for these comparisons included center-based ECE classrooms for infants and toddlers, center-based preschool classrooms, and home-based settings for infants, toddlers, and preschoolers. Overall mean scores on the CIS were comparable across INC and NSN classrooms. It is important to note these findings might be confounded by the fact that data from center- and home-based infant/toddler ECE settings were aggregated with data from center- and home-based preschool ECE settings. In addition to there being no validity evidence to suggest that the CIS measures the same constructs across INC and NSN classrooms, there have been no studies to examine the measurement invariance of the CIS across infant/toddler and preschool settings or across center- and home-based settings.

In contrast to the study conducted by Knoche et al. (2006), Wall et al. (2006) used scores from the CIS as a proxy to represent the quality of caregiver interactions experienced by individual children enrolled in center-based ECE classrooms. Data from the same types of ECE classrooms as Knoche et al. were included in this study, but were disaggregated by age and program type. For the present review, only the findings related to differences in CIS overall mean scores for children with and without special needs in center-based preschool ECE classrooms are reviewed. Comparisons were

based on data from 88 children with special needs and 311 children without special needs. Findings from this study suggested overall mean scores for the CIS were comparable for children with and without disabilities.

Although the CIS (Arnett, 1989) has been used to a much lesser extent than the ECERS-R to examine differences in the quality of caregiver interactions across INC and NSN classrooms, these data are widely available in large-scale data sets. Both of the studies reviewed represented secondary analyses of data set from relatively large-scale studies. Even if the CIS is not used commonly to examine differences in caregiver interactions across these types of programs and classrooms, it is necessary to determine whether it is appropriate to pool CIS scores across INC and NSN center-based preschool classrooms, because this practice is common in most large-scale studies of quality in ECE.

Summary

The present review of the literature included a discussion of four major topics related to the present study: (a) defining quality (b) measuring quality, (c) Early Childhood Environment Rating Scale-Revised, and (d) Arnett Caregiver Interaction Scale. The topics reviewed provided background information about how quality in ECE has been conceptualized and measured. A review of information about the development, validation, and use of the ECERS-R (Harms et al., 1998) and CIS (Arnett, 1989) provided a rationale for the present study. Information about the measurement invariance for these instruments across INC and NSN classrooms will help inform the ways in which these instruments are used in the future. Most importantly, this information will help researchers, policymakers, and practitioners interpret findings from previous research and program evaluation endeavors.

Table 2-1. Summary of Content for Selected Observation-Based Instruments Used to Measure Dimensions of Quality in ECE

Instrument	Citation	Instrument Type	Physical Environment, Materials ^a	Social-Emotional Climate ^b	Learning Environment/Opportunities	Language/Literacy	Math
Assessment Profile for Early Childhood Programs	Abbot-Shim & Sibley, 1992	Checklist	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Caregiver Interaction Scale	Arnett, 1989	Rating scale		<input type="checkbox"/>			
Classroom Assessment Scoring System	Pianta, La Paro, and Hamre, 2007	Rating scale		<input type="checkbox"/>	<input type="checkbox"/>		
Classroom Practices Inventory	Hyson, Hirsh-Pasek, and Rescorla, 1990	Rating scale		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	
Early Childhood Environment Rating Scale-Revised	Harms, Clifford, and Cryer, 1998; 2005	Rating scale	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Early Childhood Environment Rating Scale-Extension	Sylva, Siraj-Blatchford, and Taggart, 2003	Rating scale	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Early Childhood Classroom Observation Measure	Stipek and Byler, 2004	Rating scale Checklist	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Early Language and Literacy Classroom Observation	Smith, Dickinson, Sangeorge, and Anastasopoulos 2002	Rating scale Checklist	<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	
Emerging Academics Snapshot	Ritchie, Howes, Kraft-Sayre, and Weiser, 2001	Coding system			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Observation Measure of Language and Literacy Instruction	Abt Associates Inc., 2006	Coding system Checklist Rating scale			<input type="checkbox"/>	<input type="checkbox"/>	
Observation Record of the Caregiving Environment	NICHD ECCRN, 2000	Coding system		<input type="checkbox"/>			
Preschool Mathematics Inventory	National Institute for Early Education Research, 2007	Rating scale					<input type="checkbox"/>

Table 2-1 continued.

Instrument	Citation	Instrument Type	Physical Environment, Materials ^a	Social- Emotional Climate ^b	Learning Environment/ Opportunities	Language/ Literacy	Math
Preschool Program Quality Assessment	High/Scope, 2003	Rating scale	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Supports for English Language Learning Classroom Assessment	National Institute for Early Education Research, 2005	Rating scale				<input type="checkbox"/>	

Note. indicates that the instruments provides a measure of the proximal indicators associated with the content domain listed. Adapted from Snow and Van Hemel, 2008, pp. 174-177.

^aIncludes observable indicators related to safety, physical arrangement, or materials

^bIncludes observable indicators related to emotional climate, social interactions between children and adults, or support for social skill development

Table 2-2. Summary of Empirical Studies Examining Factor Structure of the ECERS

Study	N	Analyses	Extraction	Rotation	Factors
Whitebook, M., Howes, C., & Phillips, D. (1989)	313	EFA	ML	Oblique	Appropriate Caregiving ($v = 16$) Developmentally Appropriate Activity ($v = 10$)
Scarr, S., Eisenberg, M., & Deater-Deckard, K. (1994)	120	EFA	ML	Orthogonal Oblique	One-factor solution ($v = 36$)
Sakai, L. M., Whitebook., M., Wishard, A., & Howes, C. (2003)	68	PCA	PC	Orthogonal	Developmentally Appropriate Caregiving/Tone ($v = 13$) Developmentally Appropriate Activities/Materials ($v = 13$)

Note. EFA = Exploratory Factor Analysis; PCA = Principal Components Analysis; ML = Maximum Likelihood Estimation; PC = Principal Components.

Table 2-3. Summary of Empirical Studies Examining the Factor Structure of the ECERS-R

Study	N	Analyses	Extraction	Rotation	Factors
Sakai, L. M., Whitebook, M., Wishard, A., & Howes, C. (2003)	68	PCA	PC	Orthogonal	Developmentally Appropriate Caregiving/Tone ($v = 11$) Developmentally Appropriate Activities/Materials ($v = 16$)
Cassidy, D. J., Hesenes, L. L., Hegde, A., Hestenes, S., & Mims, S. (2005)	958 ^a	PCA EFA CFA	PC PF ML	Orthogonal	Activities/Materials ($v = 9$) Language/Interactions ($v = 7$)
Perlman, M., Zellman, G. L., & Vi-Nhuan, L. (2004)	326	EFA	ML	Oblique	Space/Activities/Program Structure ($v = 15$) Staff-Child Interactions ($v = 10$) Provisions for Parents and Staff ($v = 8$)
Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., & Abner, K. (2013).	1350	EFA	WLS	Oblique	Space/Activities/Program Structure ($v = 19$) Personal Care ($v = 7$) Language-Reasoning/Interaction ($v = 9$)

Note. EFA = Exploratory Factor Analysis; PCA = Principal Components Analysis; CFA = Confirmatory Factor Analysis; ML = Maximum Likelihood Estimation; PC = Principal Components; WLS = Weighted Least Squares.

^a Sample size for PCA/EFA was 486; Sample size for CFA was 472.

Table 2-4. Comparison of Factors Identified for the ECERS-R

Item	Sakai, Whitebook, Wishard, & Howes, 2003 ^a	Perlman, Zellman, & Vi-Nhuan, 2004 ^b	Cassidy, Hestenes, Hegde, Hestenes, & Mims, 2005	Clifford et al., 2005	Gordon, Fujimoto, Kaestner, Korneman, & Abner, 2013
1. Indoor Space	Developmentally Appropriate Activities & Materials	Provisions for Staff and Parents			
2. Furnishings for routine care and learning	Developmentally Appropriate Caregiving	Provisions for Staff and Parents			Activities & Materials
3. Furnishings for relaxation	Developmentally Appropriate Caregiving	Activities, Program Structure, Space	Activities & Materials		Activities & Materials
4. Room arrangement	Developmentally Appropriate Activities & Materials	Activities, Program Structure, Space		Provisions for Learning	Activities & Materials
5. Space for privacy		Activities, Program Structure, Space	Activities & Materials	Provisions for Learning	Activities & Materials
6. Child display		Activities, Program Structure, Space			Activities & Materials
7. Space for gross motor activities	Developmentally Appropriate Caregiving				Personal Care & Safety
8. Gross motor equipment	Developmentally Appropriate Activities & Materials	Activities, Program Structure, Space		Provisions for Learning	
9. Greeting and departure	Developmentally Appropriate Caregiving	Staff-Child Interactions		Teaching and interactions	Language & Interactions
10. Meals and snacks	Developmentally Appropriate Activities & Materials				Personal Care & Safety
11. Nap/rest	Developmentally Appropriate Activities & Materials				Personal Care & Safety
12. Toileting	Developmentally Appropriate Activities & Materials	Provisions for Staff and Parents			Personal Care & Safety

Table 2-4. Continued.

Item	Sakai, Whitebook, Wishard, & Howes, 2003 ^a	Perlman, Zellman, & Vi-Nhuan, 2004 ^b	Cassidy, Hestenes, Hegde, Hestenes, & Mims, 2005	Clifford et al., 2005	Gordon, Fujimoto, Kaestner, Korneman, & Abner, 2013
12. Toileting	Developmentally Appropriate Activities & Materials	Provisions for Staff and Parents			Personal Care & Safety
13. Health practices	Developmentally Appropriate Activities & Materials	Staff-Child Interactions			Personal Care & Safety
14. Safety practices	Developmentally Appropriate Activities & Materials				Personal Care & Safety
15. Books/pictures	Developmentally Appropriate Activities & Materials	Activities, Program Structure, Space	Activities & Materials		Activities & Materials; Language & Interactions
16. Encouraging communication		Staff-Child Interactions		Teaching and interactions	Activities & Materials; Language & Interactions
17. Using language to develop reasoning skills	Developmentally Appropriate Caregiving	Staff-Child Interactions	Language & Interaction	Teaching and interactions	Language & Interactions
18. Informal use of language	Developmentally Appropriate Caregiving	Staff-Child Interactions	Language & Interaction	Teaching and interactions	Language & Interactions
19. Fine motor	Developmentally Appropriate Caregiving		Activities & Materials	Provisions for Learning	Activities & Materials
20. Art			Activities & Materials	Provisions for Learning	Activities & Materials
21. Music and movement	Developmentally Appropriate Caregiving	Activities, Program Structure, Space			Activities & Materials
22. Blocks	Developmentally Appropriate Activities & Materials	Activities, Program Structure, Space	Activities & Materials	Provisions for Learning	Activities & Materials

Table 2-4. Continued.

Item	Sakai, Whitebook, Wishard, & Howes, 2003 ^a	Perlman, Zellman, & Vi-Nhuan, 2004 ^b	Cassidy, Hestenes, Hegde, Hestenes, & Mims, 2005	Clifford et al., 2005	Gordon, Fujimoto, Kaestner, Korneman, & Abner, 2013
23. Sand and water	Developmentally Appropriate Activities & Materials	Activities, Program Structure, Space		Provisions for Learning	Activities & Materials
24. Dramatic play	Developmentally Appropriate Activities & Materials		Activities & Materials	Provisions for Learning	Activities & Materials
25. Nature/Science		Activities, Program Structure, Space	Activities & Materials	Provisions for Learning	Activities & Materials
26. Math	Developmentally Appropriate Caregiving	Activities, Program Structure, Space	Activities & Materials		Activities & Materials
27. Use of TV, video, computers					
28. Promoting acceptance of diversity	Developmentally Appropriate Activities & Materials	Activities, Program Structure, Space			Activities & Materials
29. Gross motor supervision		Staff-Child Interactions		Teaching and interactions	Personal Care & Safety; Language & Interactions
30. General supervision	Developmentally Appropriate Activities & Materials	Staff-Child Interactions	Language & Interaction	Teaching and interactions	Language & Interactions
31. Discipline		Staff-Child Interactions	Language & Interaction	Teaching and interactions	Language & Interactions
32. Staff-child interactions	Developmentally Appropriate Activities & Materials	Staff-Child Interactions	Language & Interaction	Teaching and interactions	Language & Interactions
33. Interactions among children	Developmentally Appropriate Caregiving	Staff-Child Interactions	Language & Interaction	Teaching and interactions	Language & Interactions
34. Schedule	Developmentally Appropriate Activities & Materials	Activities, Program Structure, Space		Provisions for Learning	Activities & Materials

Table 2-4. Continued.

Item	Sakai, Whitebook, Wishard, & Howes, 2003 ^a	Perlman, Zellman, & Vi-Nhuan, 2004 ^b	Cassidy, Hestenes, Hegde, Hestenes, & Mims, 2005	Clifford et al., 2005	Gordon, Fujimoto, Kaestner, Korneman, & Abner, 2013
35. Free play	Developmentally Appropriate Caregiving	Activities, Program Structure, Space		Teaching and interactions; Provisions for Learning	Activities & Materials
36. Group time	Developmentally Appropriate Activities & Materials	Activities, Program Structure, Space			
37. Provisions for children with disabilities					
38. Provisions for parents		Provisions for Staff and Parents			
39. Provisions for personal needs of staff		Provisions for Staff and Parents			
40. Provisions for professional needs of staff		Provisions for Staff and Parents			
41. Staff interactions					
42. Supervision of staff		Provisions for Staff and Parents			
43. Professional growth		Provisions for Staff and Parents			

Note. Factor names are those proposed by the study authors.

^aFactors based on principal components analysis.

^bAuthors rejected the three-factor solution in favor of a one-factor solution.

Table 2-5. Comparison Groups and Interpretations from Studies Examining the Quality in ECE in Classrooms where Children with Special Needs were Enrolled

Citation	Definitions of Comparison Groups	Instrument	Unit of Analysis	No. Classrooms Sampled per Program	Unit of Interpretation
Bailey, Clifford, & Harms, 1982	Programs for children with disabilities ($n = 25$): Majority of children in the program had disabilities Programs for children without disabilities ($n = 56$): Majority of children in the program did not have disabilities	ECERS	ECE classroom	1 randomly selected	ECE program
La Paro, Sexton, & Snyder, 1998	Segregated programs ($n = 29$): Programs serving exclusively 3-to-5-year-old children with identified disabilities Inclusive programs ($n = 29$): Programs designed for typically developing, at-risk 3-to-5-year-old children, serving at least one child with an identified disability	ECERS	ECE classroom	1 randomly selected	ECE program
Buyse, Wesley, Bryant, & Gardner, 1999	Inclusive programs ($n = 62$): ECE programs enrolling at least one child with a diagnosed disability under IDEA Noninclusive programs ($n = 118$): ECE programs enrolling only children who were typically developing	ECERS	ECE classroom	1 randomly selected	ECE program

Table 2-5. Continued

Citation	Definitions of Comparison Groups	Instrument	Unit of Analysis	No. Classrooms Sampled per Program	Unit of Interpretation
Knoche, Peterson, Edwards, & Jeon, 2006	<p>Inclusive settings ($n = 58$): ECE classrooms in programs where either the ECE provider or administrator reported caring for at least one child with a verified disability during the calendar month in which the study was conducted</p> <p>Noninclusive settings ($n = 54$): ECE classrooms in programs where neither the ECE provider nor administrator reported caring for at least one child with a verified disability during the calendar month in which the study was conducted</p>	ECERS-R CIS	ECE classroom	n/a	ECE classroom
Wall, Kisker, Peterson, Carta, & Jeon, 2006	<p>Children with disabilities ($n = 91$): One or more disability indicators described by Peterson et al. (2004) were reported by a parent or EHS staff member</p> <p>Children who were typically developing ($n = 337$): None of the disability indicators described by Peterson et al. (2004) were reported by a parent or EHS staff member</p>	ECERS-R CIS	ECE classroom	n/a	Individual child

Table 2-5. Continued.

Citation	Definitions of Comparison Groups	Instrument	Unit of Analysis	No. Classrooms Sampled per Program	Unit of Interpretation
Clawson & Luze, 2008	Children with disabilities ($n = 30$): Children with individualized education plans Children without disabilities ($n = 30$): Children without individualized education plans	ECERS-R ^a	ECE classroom Individual child	n/a	Individual child
Hestenes, Cassidy, Shim, & Hegde, 2008 (Study 1)	Inclusive classrooms ($n = 458$): Classrooms in which at least one child with an identified disability was enrolled Noninclusive classrooms ($n = 854$): Classrooms in which only children who were typically developing were enrolled	ECERS-R	ECE classroom	n/a	
Hestenes, Cassidy, Shim, & Hegde, 2008 (Study 2)	Inclusive classrooms ($n = 20$): Classrooms in which at least one child with an identified disability was enrolled Noninclusive ($n = 24$): Classrooms in which only children who were typically developing were enrolled	ECERS-R	ECE classroom	n/a	
Grisham-Brown, Cox, Gravil, & Missal, 2010	Inclusive programs ($n = 33$): ECE programs serving one or more children with disabilities Noninclusive programs ($n = 33$): ECE programs serving only children without disabilities	ECERS-R	ECE classroom	1 randomly selected	

Note. Definitions of comparison groups are presented as described by study authors; number of classrooms sampled per program is only presented for studies in which inferences were made about ECE program quality.

^aIn addition to administering the instrument to characterize the quality in ECE classroom-wide, some items were adapted to characterize the quality in ECE received by individual children.

CHAPTER 3 METHODOLOGY

Research Questions

The following research questions guided the secondary analyses conducted to examine measurement invariance for the ECERS-R (Harms et al., 1998) and CIS (Arnett, 1989) INC and NSN classrooms:

1. Are the same number of latent variables measured by the ECERS-R in INC and NSN classrooms? Are the same number of latent variables measured by the CIS in INC and NSN classrooms?
2. To what extent are the items that define the latent variables measured by the ECERS-R the same in INC and NSN classrooms? To what extent are the items that define the latent variables measured by the CIS the same in INC and NSN classrooms?
3. To what extent are the factor loadings, item threshold parameters, and residual variance parameters for items on the ECERS-R invariant across INC and NSN classrooms? To what extent are the factor loadings, item threshold parameters, and residual variance parameters for items on the CIS invariant across INC and NSN classrooms?
4. To what extent are means of the latent variables measured by the ECERS-R different between INC and NSN classrooms? To what extent are means of the latent variables measured by the CIS different between INC and NSN classrooms?

Research Design

The research design for the present study involved conducting secondary validity analyses using cross-sectional data from the ECLS-B. The present study did not include primary data collection. A review of existing literature examining the factor structures for the ECERS-R and the CIS was used to generate confirmatory factor models for each instrument. Confirmatory factor analysis (CFA) is a form of latent variable modeling that is useful to test hypotheses about interrelationships between observed variables presumed to represent a common latent variable. The CFA

framework can be extended to test a number of hypotheses regarding measurement invariance of multi-item measures across populations or groups (Jöreskog, 1971; Vandeberg & Lance, 2000). In the present study, an extension of the confirmatory factor analytic framework (Bollen, 1989) was applied to examine measurement invariance of the two instruments when they were administered at the 4-year data collection wave. The comparison of interest with respect to measurement invariance was across INC and NSN classrooms. INC classrooms were defined as classrooms in which (a) both children with and without special needs were enrolled and (b) 75% or less of the enrollment constituted children with special needs at the time of data collection. NSN classrooms were defined as classrooms in which no children with special needs were enrolled at the time of data collection.

For each instrument of interest in the present study, multiple group confirmatory factor analysis (MG-CFA; Jöreskog, 1971) was used to determine whether (a) the same number of latent variables were measured; (b) the items defining these latent variables were the same across the two types of classrooms (i.e., configural invariance, Horn et al., 1983); (c) the factor loadings, item threshold parameters, and residual variance parameters were the same across the two types of classrooms (i.e., strict invariance, Meredith, 1993); and (d) the means of latent variables measured were the same across the two types of classrooms. MG-CFA is a contemporary approach for examining measurement invariance across populations or groups and is an appropriate framework for addressing the research questions in the present study (Hui & Triandis, 1985; Millsap, 2011; Steenkamp & Baumgartner, 1998; Vandeberg & Lance, 2000).

Regression analyses were also conducted to address the fourth research question focused on examining group differences in latent variables.

ECLS-B Study and Data Set

The ECLS-B was a multiple-method, multiple respondent longitudinal study of children's home and educational experiences from birth through kindergarten entry. The ECLS-B data set contains data from a nationally representative sample of 14,000 children born in the United States in 2001 (Snow et al., 2007). There were five waves of data collection, during which researchers collected data on children's health, development, care, and education. Data collection waves occurred when the children were approximately 9 months old (2001), approximately 2 years old (2003), preschool age (approximately 4 years old, 2005), when the children were 5 to 6 years old (kindergarten 2006), and when children were 6 to 7 years old (kindergarten 2007).

More information about the timing of the data collection waves is provided in Chapter 1.

Data sources for the ECLS-B included computer-assisted parent interviews, self-administered parent questionnaires, interviews with ECE providers, interviews with ECE center directors, observations of children's ECE settings, and teacher questionnaires.

ECLS-B Sampling Procedures

The ECLS-B used a clustered, list frame sampling design to obtain a nationally representative sample of children born in 2001. The sampling frame consisted of 14,000 births drawn from registered births in the National Center for Health Statistics vital statistics (Snow et al., 2007). Births were sampled from 96 individual counties or groups of adjacent counties (i.e., primary sampling units). Thirty-six sampling strata were generated for three different analytical domains: (1) race, (2) birth weight, and (3) and birth plurality. Five analytic groups were over-sampled to ensure enough cases for

precision weighting to make the selected cases a nationally representative sample: (1) a supplemental sample of American Indian/Alaska Natives, (2) Chinese, (3) other Asian/Pacific Islander, (4) moderately low birth weight, which was defined as infants weighing at least 1500 g and less than 2500 g at birth, and (5) very low birth weight, which was defined as infants weighing less than 1500 g at birth. An additional 18 primary sampling units were selected from a supplemental frame of areas with high proportions of American Indian/Alaska Native births to support the American Indian/Alaska Native oversample (Snow et al., 2007). Children were excluded from the sample if (a) the mother was less than 15 years of age at the time of child's birth; (b) the child died before the 9-month assessment; or (c) the child was adopted before the 9-month assessment. In subsequent waves of data collection, children who were previously in the sample were excluded if they had died or moved permanently out of the United States before the wave of data collection. Data for the ECLS-B were released in waves according to the data collection schedule. At the end of each wave of data collection, a new data file that included data from the recently completed data collection wave and all prior data collections waves was prepared and made available as a restricted-use data set.

Sampling Procedures for Preschool Child Care Observations

The ECLS-B data set includes data collected during observations of children's primary non-parental care arrangements for a subsample of the children in the ECLS-B study at the 2-year and preschool (4-year) waves of data collections. Data from the preschool wave of child care observations (CCO) were used in the present study. Cases were eligible for a CCO at the preschool data collection wave if (a) a completed ECE provider telephone interview in English or Spanish was available; (b) the center

director or ECE provider gave verbal permission for the observation; (c) the child was in a non-parental care arrangement for more than 10 hour per week; (d) the child was usually in the care arrangement for at least a 2 ½ hour block of time; (e) the child was usually awake during the time he/she was in the care arrangement; (f) the child did not live in Alaska or Hawaii at the time of the 9-month and 2-year data collection waves; (g) the child was not part of the supplemental American Indian/Alaska Native sample; and (h) the parent provided a signed permission form for the observation (Snow et al., 2007).

Preschool CCO sampling rates were designed to obtain a targeted number of CCOs across home-based ECE programs, Head Start programs, and other center-based programs. In addition, sampling targets were set for children across three poverty level groups: (a) less than 100% of poverty level, (b) 100% to 150% of poverty level, and (c) greater than 150% of poverty level. The aim was to obtain sample sizes sufficient to detect effect sizes of 0.30 for the following pairwise tests within the overall CCO sample and within the group of children who lived at less than or equal to 150% poverty level: (a) home-based versus Head Start settings, (b) home-based care settings versus center-based programs other than Head start, and (c) Head Start programs versus other center-based (Snow et al., 2007).

The final preschool child care observation subsample included 650 children who had a CCO at age 2 and met the inclusion criteria for a CCO at the preschool data collection wave and 1100 children who did not have a CCO at age 2 but who were eligible for a CCO at the preschool data collection wave. A total of 1750 CCOs were completed in the preschool data collection wave, constituting approximately 24% of

cases in the total ECLS-B sample for which non-parental care was reported. Of the total sample of CCOs, 1400 were conducted in center-based ECE programs and 350 were conducted in home-based programs. Center-based programs included public prekindergarten, private prekindergarten, Head Start, child care centers, preschools or nursery schools, or other program settings that were not home-based. The exclusions noted previously represent restrictions in the population of inference for the preschool CCO subsample. For the present study, the population of inference is limited to preschool-aged children who were in center-based care for 10 or more hours per week in the continental United States in 2005-2006. It does not include children who received non-parental care in settings where the primary language spoken was not English or Spanish.

ECLS-B Response Rates and Sampling Weights

The response rates and weighting procedures for the total ECLS-B sample and for the subsample of children whose preschool ECE classrooms were observed are relevant for the present study. Weighted and unweighted response rates for each of these samples are provided. Unweighted response rates reflect response rates given eligibility for a particular data collection component, whereas weighted response rates reflect the proportion of the sample that responded, after adjusting for sampling procedures. Unweighted response rates were calculated by dividing the number of completed cases by the total number of eligible cases sampled. To calculate weighted response rates, each sampled case was first multiplied by the inverse of its probability for selection for the study and then entered into the same equation used to calculate unweighted response rates (Snow et al., 2007). A description of the procedures used to construct sampling weights is presented below.

Response rates for total ECLS-B sample

Over 14,000 births from the 2001 sampling frame were fielded in the 9-month wave of ECLS-B data collection. The response rates presented represent the final status of the parent interview for the data collection wave relative to the total sample of eligible cases fielded in the 9-month data collection wave. The weighted and unweighted response rates for parent interviews were 76% ($N = 10,700$) and 74%, respectively (Nord et al., 2004). All cases for which a parent interview was completed at the 9-month data collection wave were fielded in the 2-year wave of data collection, which resulted in an unweighted parent interview completion rate of 70% ($N = 9800$) and a weighted completion rate of 68% (Nord et al., 2006). In the preschool wave of data collection, all cases with a completed parent interview at the 2-year data collection wave and cases that were part of the American/Indian Alaska Native supplement were fielded. A total of the 9900 cases were fielded, with an unweighted parent interview response rate of 64% ($N = 8950$) and a weighted completion rate of 62% (Snow et al., 2007).

Response rates for CCO subsample

Response rates for the preschool CCO subsample were calculated based on the number of CCO completions relative to the number of cases for which there was a completed parent interview. Of the 8950 children for which there was a completed parent interview at the preschool data collection wave, 7300 were reported to have received non-parental care (Snow et al., 2007). ECE interviews were fielded for 6650 of these cases, with unweighted and weighted response rates of 67% ($n = 6600$). A total of 4050 cases were fielded for the CCO; however, the response rate for CCO was much lower than other data collection components due to the additional levels of permissions

required to conduct observations. The unweighted CCO response rate was 19% ($n = 1800$), which includes some cases for which there were completed center director interviews but no accompanying classroom observations (Snow et al., 2007). The weighted CCO response rate was 21%.

Calculation of sampling weights

Sampling weights for the ECLS-B data were constructed in three steps. ECLS-B researchers constructed base weights from overall selection probabilities in the initial sampling frame, which were then adjusted for survey non-response at each wave of data collection and across data collection waves (Snow et al., 2007). In a final step, raking, a procedure to adjust the sums of weights to a known or estimated total, was used to improve precision of survey estimates. Sampling weights were calculated based on response rates for the data collection components at each wave of data collection. Theoretically, a sampling weight should be obtained for each component or combination of components within and across the waves from which data were drawn. Although it is preferable theoretically to obtain a sampling weight that adjusts for selection probability and non-response for each component of interest within and across every wave of data collection, the process for obtaining such weights for every possible combination of data collection components was described as not feasible or practical (Snow et al., 2007). The sampling weights provided for each wave of data collection represent weights for data collection component combinations determined to likely be of high analytical interest. The ECLS-B data set contains four sets of weights for the 9-month wave of data collection, and 10 sets of weights each for the 2-year and 4-year data collection waves.

ECLS-B Instrumentation

The ECLS-B restricted-use data file includes data from a number of data sources. For each wave of data collection, ECLS-B researchers collected data via an in-person, computer assisted parent interview, direct child assessments, and a self-administered questionnaire for fathers. Data from parent interviews provide information about family and child demographics, the family's use of subsidies, the child's health and development, home environment, and family routines. Direct child assessments were used to collect information about the child's cognitive development (i.e., color knowledge, language, literacy, math), physical development (i.e., fine motor, gross motor), and socio-emotional development (i.e., attachment, attentiveness, interest, affect, social engagement). In addition to the parent interviews and direct child assessments conducted across all five data collection waves, the 9-month data collection wave also included data collected from a review of children's birth certificates. At the 2-year and 4-year data collection waves, interviews with children's primary non-parental care providers were conducted. For children who were enrolled in home-based ECE programs, an interview was conducted with the primary ECE provider to obtain information about the home-based program, the ECE provider's educational background and experience in early childhood education, and the ECLS-B focal child's experiences in the home-based care setting. For children who were enrolled in center-based ECE programs, the center director completed a questionnaire that provided information about the program and the center director's educational background and experience in early childhood education.

Following the center director interview, an interview was conducted with the ECLS-B focal child's primary ECE provider to obtain information about the ECE provider's educational background and experience in early childhood education and the ECLS-B focal child's experience in the center-based ECE program (ECLS-B, 2003b, 2005b). The ECLS-B data set also includes data from CCOs that were conducted for a subsample of children who were enrolled in non-parental care for more than 10 hours per week at the 2- or 4-year waves of data collection. The data file includes item-level data from the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989), which were used to characterize dimensions of quality in the ECE classrooms in which ECLS-B children were enrolled. At the kindergarten waves of data collection, interviews were conducted with children's ECE providers and ECE providers that provided after-school care (i.e., wrap-around care) when applicable. Children's kindergarten teachers completed questionnaires, which provided demographic information about themselves, teacher ratings of children's development, and characteristics of the classroom in which the ECLS-B focal children were enrolled.

Data for the present study were drawn from the subsample of children whose center-based ECE classrooms were observed in the 4-year follow up of the ECLS-B. A number of data sources from the 9-month, 2-year, and preschool waves of ECLB data collection were used in the present study. Information from birth certificates and parent interviews conducted at the 9-month, 2-year, and 4-year waves of data collection was used to create composite variables characterizing whether the ECLS-B focal child had special needs. Information from the preschool parent interview was used to provide demographic information about the child and family at the time ECERS-R (Harms et al.,

1998) and CIS (Arnett, 1989) data were collected. Information from the preschool ECE provider interview was used to characterize classrooms as either INC or NSN classrooms as defined in the present study. Item-level data from the ECERS-R and the CIS were used to conduct multiple-group confirmatory factor analyses. Derived composite scores were generated from item-level ECERS-R and CIS data and used to examine differences in derived composite scores across INC and NSN classrooms. A description of the data sources used in the present study, including the two measurement instruments of primary interest for the study, is provided below.

Birth Certificate Data

The first wave of data collection for the ECLS-B included a review of birth certificate records. Birth certificate data in the data set included a number of variables related to (a) the child's date of birth, (b) characteristics of the child's place of birth, (c) demographic characteristics of the child's mother and father, (d) maternal/neonatal history, (e) gestation, (f) child gender, (g) plurality, (h) the child's health status at birth, (i) method of delivery, (k) risk factors for pregnancy, (l) obstetric procedures, (m) complications of labor or delivery, (n) abnormal conditions of the child at birth, and (o) congenital anomalies of the child. A detailed list of the variables included under each of these categories is provided in the ECLS-B preschool user's manual (Snow et al. 2007, pp. 375-379).

Parent Interviews

Computer-assisted parent interviews were conducted at every wave of ECLS-B data collection. To the extent possible, the interview respondent was the child's biological mother and the information gathered was relatively consistent across waves. The most common reason for inconsistencies in parent interview questions across

waves was a change in the extent to which questions were developmentally appropriate. For example, the preschool parent interview includes questions about school readiness that were not asked in the parent interviews at the 9-month or 2-year data collection waves, because children were not preparing to enter school at those time points. The parent interviews consisted of questions about the following major areas: (a) family structure, (b) child development, (c) home environment, (d) parenting behavior and attitudes, (e) child care arrangements, (f) child health, (g) family health, (h) respondent's marriages and relationships, (i) community and social support, (j) respondent information, (k) spouse/partner information, (l) non-resident father information, (m) welfare and other public assistance, (n) household income and assets, and (o) neighborhood quality and safety. The battery of questions asked as part of the parent interviews are available via the ECLS-B website (IES, NCES, n. d.)

ECE Provider Interviews

Interviews were conducted with the ECLS-B focal child's primary preschool ECE provider (i.e., the ECE provider with whom the child spent the most time in non-parental care) when applicable and when consent could be obtained from the ECE provider. The topics covered in the ECE provider interview were: (a) type of care, (b) center information, (c) staffing, (d) center services, (e) transition to the caregiver/teacher, (f) care of the focal child, (g) other children in care, (h) child development, (i) caregiver-child relationship, (j) parental involvement, (k) caregiver attitudes and beliefs, (l) learning environment, (m) curriculum and activities, (n) caregiver background, (o) professional development, (p) caregiver health, (q) caregiver income. The battery of questions asked as part of the ECE interviews are available via the ECLS-B website (IES, NCES, n. d.)

Child Care Observations

The primary data sources for the present study were from instruments completed as part of observations conducted in the primary center-based ECE classrooms in which ECLS-B focal children were enrolled. The purpose of these observations was to provide information about the global, structural, and process quality of the center-based ECE classrooms in which ECLS-B focal children spent more than 10 hours per week. The child care observations (CCOs) were conducted by staff from the Research Triangle Institute from October 2005 to June 2006 and included the administration of the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989). Each observation took approximately 3 ½ hours to complete. The majority of this time was spent observing and scoring the ECERS-R, which requires a block of at least 3 hours for an external assessor to administer (Harms et al., 1998). The CIS was scored at the end of the CCO (Snow et al., 2007). CCO observers met at least two of the following three criteria: (a) had experience working with young children or working in child care, (b) held a bachelor's degree in early childhood or a related field, (c) had experience working on research studies for which observational data were collected or that involved child care or schools (Snow et al., 2007). CCO field observers underwent extensive training for ECLS-B certification. A description of each instrument, the certification procedures for each instrument, and the data from each instrument that are available in the ECLS-B data file is provided.

ECERS-R

The ECERS-R is an observation, judgement-based rating scale consisting of 43 items designed to assess dimensions of quality in group programs providing ECE services to young children between the ages of 2 and 5 years (Harms et al., 1998).

Items on the ECERS-R are scored on a scale from 1 (*inadequate*) to 7 (*excellent*).

Each item is scored based on the presence or absence of quality indicators related to the arrangement of the physical environment, provisions of materials or activities, supervision of children, and interactions between individuals in the classroom.

Indicators are anchored at four scale ratings: 1 (*inadequate*), 3 (*minimal*), 5 (*good*), and 7 (*excellent*). A rating of 1 for an item generally indicates that there is no provision in either the physical environment or available materials for the area of quality measured by the item. A rating of 3 indicates the provision of minimal materials, space, or supervision. A rating of 5 indicates adequate materials and space set aside for activities as well as adequate supervision of children. A rating of 7 indicates evidence of teacher planning for individual child differences and encouragement of children's independence within activities in addition to adequate space, materials, and supervision (Harms & Clifford, 1982). Mid-point ratings of 2, 4, and 6 are scored when all of the indicators anchored at the lower anchor are met, but less than half of the indicators at the higher anchor are met.

In a field test with 21 classrooms, the percentage agreement across the 470 observable indicators that comprise the 43 items on the ECERS-R was 86.1% (Harms et al., 1998). There was no item for which the indicator level agreement was below 70%. Exact item-level agreement was 43%, and percentage agreement within 1 scale point was 71%. For the overall ECERS-R score, Pearson product moment and Spearman rank order correlations were .921 and .865, respectively (Harms et al., 1998).

The developers of the ECERS-R (Harms et al., 1998) proposed seven subscales of quality measured by the ECERS-R: (a) Space and Furnishings ($v = 8$), (b) Personal

Care Routines ($v = 6$), (c) Language and Reasoning ($v = 4$), (d) Activities ($v = 10$), (e) Interaction ($v = 5$), (f) Program Structure ($v = 4$), and (g) Parents and Staff ($v = 6$).

Table A-1 shows the subscales and corresponding items from the ECERS-R. Subscale scores are obtained by averaging the item scores relevant to each subscale. An overall score is obtained by averaging scores across all items.

In a field test conducted by the developers of the ECERS-R, internal consistency score reliabilities for subscale scores ranged from .71 to .88 and the internal consistency score reliability for the total score was .92 (Harms et al., 1998).

Subsequent studies in which the psychometric properties of the ECERS-R have been examined have reported similar internal consistency score reliabilities (e.g., Cassidy, Hestenes, Hegde, 2005; Perlman et al., 2004; Snow et al., 2007).

As part of the ECLS-B, the first six subscales of the ECERS-R (Harms et al., 1998) were administered in 1400 center-based preschool ECE classrooms. The seventh subscale (i.e., Parents and Staff) was not administered because the items comprising this subscale could not be observed directly, and information relevant to these items could be obtained via interview (Snow et al., 2007). Information regarding the procedures followed to certify CCO field observers to administer the ECERS-R and the ECERS-R data available in the ECLS-B data file is provided below.

ECERS-R certification

CCO observers were trained to administer the ECERS-R (Harms et al., 1998) by trainers who had learned to administer the ECERS-R at the institute in which it was developed and who had subsequently conducted ECERS-R trainings for this institute (Snow et al., 2007). Before achieving certification to administer the ECERS-R for the ECLS-B, observers were required to participate in training procedures, which included 5

days of classroom training and practice observations in center-based preschool classrooms. The classroom training included approximately 6 hours total of didactic teaching related to the ECERS-R, in addition to opportunities to practice scoring ECERS-R items from video. When classroom training was complete, small groups of one CCO trainer and three CCO field observers spent two days scoring the ECERS-R in center-based preschool classrooms. Within a small group, the trainer and field observers conducted simultaneous observations in which they scored the CCO instruments without consulting each other. When scoring was completed, each small group held a group discussion, during which consensus scores for each item were obtained. The consensus scores that were obtained were used as the standard for calculating interrater agreement. CCO field observers were required to achieve 80% interrater agreement within 1 scale point (on a 7-point scale). Percent agreement was calculated by dividing the number of item-level agreements by the total number of item-level agreements and disagreements and multiplying by 100. If agreement standards were not met in the initial field practice days, additional practice days were arranged.

The majority of CCO field observers achieved certification during the initial two field practices days (Snow et al., 2007). The average percent agreement for these observers was 92% (*range* = 81 – 100). For CCO observers who were certified after four practice observations, the average percent agreement was 95% (*range* = 85 – 100). There were some raters who did not achieve agreement standards during practice observations, but who received provisional certification based on positive reports from the CCO trainers. Average percent agreement for these observers was 77% (*range* = 76 – 78).

ECERS-R data provided in the ECLS-B data file

The ECLS-B data file contains continuous subscale composite variables from the six ECERS-R subscales that were administered during the 4-year follow-up of the study: (a) Space and Furnishings ($v = 8$), (b) Personal Care Routines ($v = 6$), (c) Language and Reasoning ($v = 4$), (d) Activities ($v = 10$), (e) Program Structure ($v = 4$), and (f) Interactions ($v = 5$). Subscale scores were obtained by taking the average of item scores within the subscale. Table A-1 displays the ECERS-R items that were reported by the instrument developers to compose each subscale and were used to calculate subscale scores. In addition to subscale composite scores, a total composite score for the ECERS-R is provided in the ECLS-B data file. The total score provided is a continuous variable derived from averaging item scores for 36 of the 37 items that were administered. The item pertaining to provisions for children with disabilities (i.e., item 37) was not included in the calculation of the overall score, because it was only administered in classrooms in which the ECE provider reported at least one child with an identified disability was enrolled. Internal consistency score reliabilities for overall mean score and for each of the subscale scores for the analytic sample of classrooms used in the present study were similar to those reported in field tests of the ECERS-R (Harms et al., 1998), ranging from .72 to .94 for the total sample (see table 3-1). Internal consistency score reliabilities for INC and NSN classrooms were similar, ranging from .71 to .94 and .73 to .95, respectively.

Of primary interest for the present study were the item-level ECERS-R (Harms et al., 1998) data provided in the ECLS-B data file for INC and NSN classrooms. Each item-level score is an ordinal categorical variable with a value ranging from 1 to 7. Item-level data were available for each of the 37 ECERS-R items that were administered

during the preschool CCO. For the present study, item-level scores for 34 ECERS-R items were used in multiple-group confirmatory factor analyses to examine measurement invariance of the ECERS-R across INC and NSN classrooms that were observed in the 4-year follow up of the ECLS-B. Three items (i.e., nap/rest, use of TV/video/computers, provisions for children with disabilities) were excluded from the analyses due to missing data patterns that were related to the fact that the items could receive a score of “not applicable.” In addition to using item-level data to conduct multiple-group confirmatory factor analyses, item-level data were used to generate derived composite scores for the latent variables derived in the multiple group confirmatory factor analyses for the purpose of examining differences in the derived composite scores across INC and NSN classrooms.

Caregiver Interaction Scale

The Caregiver Interaction Scale (CIS; Arnett, 1986; 1989) was administered in 1400 center-based ECE programs and 350 home-based ECE programs during the 4-year follow up of the ECLS-B. The CIS is a judgment-based observational rating scale that was designed to provide a global quality rating of caregiver interactions. It is composed of 26 items that represent observable caregiver behaviors. Items are scored on a scale from 1 (*not at all*) to 4 (*very much*). This scale signifies the extent to which the caregiver engaged in the behavior described on the CIS. As shown in Table A-2, items on the CIS are either positively (e.g., speaks warmly to children) or negatively (e.g., seems critical of the children) oriented. The developer of the CIS proposed four subscales describing the quality of caregiver interactions measured by the CIS: (a) Sensitivity, (b) Harshness, (c) Detachment, and (d) Permissiveness. As shown in Table A-2, the Harshness and Detachment subscales are comprised mostly of negatively

oriented items. Before calculating composite scores, negatively oriented items are typically reverse scored so that a higher subscale indicates more positive interactions and a lower score indicates less positive interactions. Subscale scores on the CIS typically are calculated by averaging item-level scores associated with a subscale. The total CIS score is calculated by averaging scores across all items.

Slight adaptations were made to the CIS (Arnett, 1986; 1989) for use in the ECLS-B. A comparison of the items from the original instrument and the instrument used in the ECLS-B is provided in Table A-2. In general, the adaptations made to CIS items included the addition of examples to facilitate scoring. The adapted CIS items are similar to those used in other large-scale studies, such as the Head Start Family and Child Experiences Survey and the EHSRE (Snow et al., 2007).

Caregiver Interaction Scale certification

Procedures for certifying CCO field observers to administer the CIS (Arnett, 1986; 1989) for the ECLS-B were similar to those described previously for the ECERS-R (Harms et al., 1998). The 5-day classroom training described previously included approximately 3 ½ hours total of didactic teaching related to the CIS, in addition to opportunities to practice scoring CIS items from video. When classroom training was complete, small groups of one CCO trainer and three CCO field observers practiced scoring the CIS in center-based preschool classrooms and home-based preschool classrooms. Two practice days were devoted to practice scoring in center-based preschool classrooms and two days were devoted to practice scoring in home-based preschool classrooms. Within a small group, the trainer and field observers conducted simultaneous observations in which they scored the CCO instruments without consulting each other. When scoring was completed, each small group held a group

discussion, during which consensus scores for each item were obtained. The consensus scores that were obtained were used as the standard for calculating interrater agreement. CCO field observers were required to achieve 80% interrater agreement within 1 scale point (on a 4-point scale). Percent agreement was calculated by dividing the number of item-level agreements by the total number of item-level agreements and disagreements and multiplying by 100. If agreement standards were not met in the initial field practice days, additional practice days were arranged.

The majority of CCO observers were certified during the initial practice observations (Snow et al., 2007). The average percent agreement for these observers was 98% (*range* = 85 – 100). For observers who were certified after four practice observations, the average percent agreement was 94% (*range* = 88 – 100). None of the CCO field observers required provisional certification.

Caregiver Interaction Scale data provided in the ECLS-B data file

The ECLS-B data file contains continuous composite score variables for the four subscales of the CIS (i.e., Sensitivity, Harshness, Detachment, Permissiveness) and a total composite score. ECLS-B researchers conducted two data transformations before calculating subscale and total score composites. First, each CIS item score was recoded on a 0-3 scale. Second, negatively oriented items (e.g., speaks with irritation or hostility to the children) were reverse coded so that a higher score indicated a higher quality interaction and a lower score indicated a lower quality interaction. The composite subscale scores provided in the ECLS-B data file were calculated by summing the item scores associated with the subscale. Subscale composites were only provided if one or fewer of the items in each of the four subscales were missing. The total score composite provided in the data file was calculated by summing all item-level

scores. Internal consistency reliability scores for the total CIS score and subscale scores for the analytical sample classrooms used in the present study ranged from .59 to .95. Internal consistency score reliability was similar across INC and NSN classrooms, ranging from .61 to .95 and .54 to .94, respectively (see Table 3-1).

Of primary interest for the present study were the CIS (1989) item-level scores from observations INC and NSN classrooms center-based preschool classrooms. The CIS item-level scores provided in the ECLS-B data file are ordinal categorical variables that range from 1 (*not at all*) to 4 (*very much*) rating scale. Although the composite scores provided in the data file reflect reverse-coded item level data for negatively oriented items, the item-level variables provided in the data file were not reverse-coded. For the present study, the item-level data provided in the ECLS-B data file were reverse-coded (when necessary) and used in multiple-group confirmatory factor analyses to examine measurement invariance for the CIS across INC and NSN classrooms. In addition, item-level data were used to generate derived composite scores for the latent variables derived in the multiple group confirmatory factor analyses for the purpose of examining differences in derived composite scores across INC and NSN classrooms.

Definition of Key Variables

Before statistical analyses could be conducted for the present study, it was necessary to define two key variables. First, it was necessary to define “children with special needs.” For this study, there were two groups of children with special needs identified: ECLS-B focal children with special needs and non-focal children who had special needs and who were enrolled in the same classrooms as the ECLS-B focal children. After determining criteria for defining children with special needs, it was

necessary to define INC and NSN classrooms. The definitions and data sources used to define these key variables are described below.

Children with Special Needs

Two groups of children with special needs were identified in the present study. First, criteria were determined to identify whether ECLS-B focal children had special needs. Second, criteria were determined to identify whether other children enrolled in the same classrooms as ECLS-B focal children had special needs. The types of data available for these two groups of children differed, therefore, the data sources and criteria for identifying children with special needs differed slightly depending on whether the child was an ECLS-B focal child.

ECLS-B focal children with special needs

Information from the 9-month, 2-year, and preschool parent interviews and from birth certificates was used to determine whether the ECLS-B focal child had special needs. At each of these three waves of data collection, parents were asked a series of questions about the child's health and development. A list of the questions used in the present study to identify whether the ECLS-B focal child had special needs is provided Appendix B. One question pertained to whether the child was receiving early intervention or special education services. A second question asked if a doctor had told the parent the child had a medical condition since the time of the last interview. For this question, the interviewer provided the respondent with a list of specific conditions (e.g., spina bifida, Down Syndrome, epilepsy). In the preschool parent interview, an additional series of questions was asked related to whether the child had ever been evaluated and received a diagnosis of a problem for a number of conditions (e.g., difficulty hearing or understanding communication, difficulty paying attention, difficulty

moving or using limbs). For the present study, if the parent indicated the child (a) received early intervention or special education services, (b) received a diagnosis of a medical condition from a doctor, or (c) was evaluated and received a diagnosis of a problem from a professional at the 9-month, 2-year, or preschool data collection waves, the child was characterized as having special needs. In addition, children who were identified from birth certificate data as having very low birth weight (i.e., < 1500 g) were characterized as having special needs.

The procedures for identifying children with special needs in the present study are aligned with those described by Peterson et al. (2004). Peterson et al. defined children with special needs as children who met at least one of the following criteria: (a) received early intervention or preschool special education services, (b) had diagnosed conditions that would make them eligible for early intervention or preschool special education services or that were likely to be associated with developmental delays (e.g., sensory impairment, chromosomal abnormality), (c) had suspected delays (e.g., difficulty using arms or legs), as reported by parents or professionals, or (d) had biological risks (e.g., low birth weight). These criteria were proposed using data from similar questions posed in family interviews used in the Head Start FACES study (OPRE, 1997-2013; Peterson et al., 2004)

Non-focal children with special needs

The primary focus of the ECLS-B was the focal children who were followed from birth through kindergarten entry, therefore, there was limited data available regarding characteristics of other children enrolled in the same classroom as the ECLS-B focal child. Given this limitation, the criterion for identifying non-focal children as children with special needs differed slightly than the criteria used for focal children. One question

from the ECE provider interview was used as the basis for this determination. The ECE provider was asked to report how many children with special needs were typically provided care at the same time as the focal child. Children with special needs were defined in the ECE provider interview as children who had diagnosed disabilities, medical problems, or emotional problems (Institute of Education Sciences [IES], National Center for Education Statistics [NCES], n. d.). The ECLS-B definition for special needs was generally aligned with the criteria used in the present study to identify whether ECLS-B focal children had special needs.

INC and NSN classrooms

Information about whether the ECLS-B focal child had special needs, whether there were other, non-focal children in the classroom with special needs, and total classroom size was used to calculate the proportion of children with special needs in each classroom. Classrooms in which children with and without special needs were enrolled and for which 75% or less of the enrollment consisted of children with special needs at the time of the ECE provider interview were characterized as INC classrooms. Classrooms in which only children without special needs were enrolled at the time of the ECE provider interview were characterized as NSN classrooms. Classrooms in which more than 75% of the enrollment consisted of children with special needs were excluded from the present analyses because they were hypothesized to be qualitatively different than the other two types of classrooms and the sample size was not sufficient to include in the analyses conducted to examine measurement invariance in the present study. The procedures used to calculate the proportion of children with special needs in each classroom are described in the Analytical Procedures section of this chapter.

Analytic Sample for the Present Study

The analytic sample for the present study consisted of the data from a subsample of children who (a) participated in the ECLS-B study, (b) had a CCO conducted at the 4-year data collection wave, and (c) were enrolled in either an INC or NSN classroom. A total of 1350 classrooms were included in the analytic sample for the present study. Of the 1350 classrooms included, 900 were INC classrooms and 450 were NSN classrooms. The average age of ECLS-B focal children enrolled in INC classrooms was 56 months (*range* 46 - 66) at the time of the CCO. The average age of ECLS-B focal children who were enrolled in NSN classrooms was 56 months (*range* 46 - 65) at the time of the CCO. Table 3-2 provides additional descriptive information about the ECLS-B focal children whose INC or NSN classrooms were observed. The characteristics of ECLS-B focal children were similar across INC and NSN classrooms, with the exception of poverty status and SES. A higher percentage of ECLS-B focal children who were enrolled in INC classrooms lived below the poverty threshold and were in lower SES quintiles.

The two types of classrooms included in the present analytic sample had similar group sizes and adult-child ratios; however, there was a broader range of group size for NSN classrooms. The mean group size for INC classrooms was 15 children, with a range of 2 to 31 children. The mean group size for NSN classrooms was also 15, with a range of 1 to 51 children. Average adult-child ratios for INC and NSN classrooms were 7 (*range* = 1 - 22) and 8 (*range* = 1 - 22), respectively. The mean proportion of children with special needs in INC classrooms was 19% (*range* 3% - 75%). The mean proportion of children with special needs includes ECLS-B focal children with special needs as well as non-focal children were reported to have special needs by their

teachers. The majority of ECLS-B focal children who were enrolled in INC classrooms were identified as having special needs either through a medical diagnosis or through an evaluation by a professional (see Table 3-2). There was no information available the types of special needs for non-focal children who were enrolled in INC classrooms.

The classrooms included in the present analytic sample were located in one of the following types of center-based programs: (a) public prekindergarten, (b) private prekindergarten, (c) child care center, (d) Head Start center, (e) preschool/nursery school, or (f) other center-based program. A higher percentage of INC classrooms were located in public prekindergarten or Head Start programs compared to NSN classrooms (see Table 3-3). In addition, a higher percentage of NSN classrooms were located in private prekindergarten, child care, or preschool/nursery school programs compared to INC classrooms. Table 3-4 displays the weighted means of the ECERS-R (Harms et al., 1998) and CIS (Arnett, 1989) subscales proposed by the instrument developers for the total combined analytic sample for the present study ($N = 1350$), for INC classrooms ($n = 900$), and for NSN classrooms ($n = 450$). In general, INC classrooms had higher scores on the published ECERS-R subscales than NSN classrooms. Scores for the published CIS subscales were similar across groups. Given the lack of empirical evidence validating published subscales for these instruments, no statistical comparisons of scores on published subscales were made.

Analytic Procedures

In this section of the dissertation, the analytic procedures carried out to address the proposed research questions are described. Procedures used to prepare data for analysis, weighting procedures, and the analyses conducted to examine each research question are presented.

Data File Preparation

Initial data files for the present study were constructed using the ECLS-B Electronic Code Book (ECB). The ECB contains information about each variable included in the ECLS-B data file, including the variable name, response categories, and the frequencies for each response category. The ECB allows researchers to search for and select specific variables from the ECLS-B data. Select variables can be compiled and saved in a taglist, which is used to generate syntax that extracts only the selected variables from the full ECLS-B data file.

For the present study, the ECB was first used to generate an SPSS file with the variables of interest for all ECLS-B cases. The ECERS-R total score variable was used to identify which cases from the full data file had a preschool CCO in a center-based setting, and these cases were extracted from the file to create the data file that was used in subsequent analyses for the present study. SPSS syntax was written to recode missing data. Given that the data used in the present study were cross-sectional and drawn from a limited number of sources for a subsample of cases in the ECLS-B data file, there were very few cases with missing data. The syntax used to recode missing data is shown in Appendix C.

After missing data were recoded, a number of composite variables were constructed using SPSS syntax. One variable was constructed to identify whether the ECLS-B focal child had special needs (see Appendix B for a list of variables and responses used to construct this composite). Three composite variables were constructed to identify whether each of the ECE classrooms observed during the preschool CCO was: (a) a classroom in which the ECLS-B focal child had special needs; (b) a classroom in which the ECLS-B focal child did not have special needs, but

the ECE provider reported that both children with and without special needs were enrolled; or (c) a classroom in which none of the children enrolled, including the focal child, had special needs. A separate data set with all the ECLS-B variables of interest was constructed for each of the three types of classrooms. Within each data set, SPSS syntax was used to construct a composite variable defining the proportion of children with special needs using information about (a) whether the ECLS-B focal child had special needs, which was derived from the first composite variable constructed for the present study (b) the number of other children in care who had special needs, which was derived from a question asked in the ECE provider interview; and (c) the total number of children enrolled in the classroom, which was derived from a question asked in the ECE provider interview (see syntax in Appendix C). As shown in Appendix C, the equations used to identify the number of special needs and the proportion of children with special needs was slightly different in each of the three data sets due to the nature of the questions from the ECE provider interview that were used to construct this variable.

A final composite variable was constructed that identified whether the center-based ECE classroom observed for the preschool CCO was an INCC or a CNSN. Following the construction of composite variables describing the proportion of children with special needs and the classroom type, the three data sets were merged to create one master data file that consisted of all 1350 cases and all the ECLS-B variables of interest for the present study.

The master data file was imported to SAS™ 9.3 to conduct descriptive analyses. Two additional data files were created and imported into Mplus 7 (Muthén & Muthén,

1998-2012) to conduct multiple group confirmatory factor analyses. First, a data set containing only the classroom type variable, ECERS-R (Harms et al., 1998) item-level variables, and weights was created. A similar data set for the CIS (Arnett, 1989) was also created. Within the CIS data set, negatively oriented items were recoded so that a higher score represented a higher quality interaction. After confirming the internal structure for the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989), derived composite scores were constructed in SPSS by averaging the item-level scores for the items associated with each latent variable indicated in the measurement model for each instrument. The data files with the item-level data, weights, and derived composites were exported to SAS™ 9.3 to conduct group comparisons of derived composites across INC and NSN classrooms and to calculate internal consistency score reliabilities for the composite variable subscales.

Weighting

The ECLS-B data file includes 10 sets of weights for the preschool wave of data collection. Weights were calculated to adjust for overall sampling probability and for survey nonresponse patterns within and across data collection waves. The 10 sets of weights provided for the preschool wave of data collection represent combinations of data components that were likely to be of analytical interest to researchers; the weights provided do not represent every possible combination of data components (Snow et al., 2007). The weights used in the present study adjusted for stratification of the sample, cluster sampling, and overall sampling probability in the base year of data collection as well as survey non-response patterns for the following data components: (a) parent interviews from the 9-month, 2-year, and preschool data collection waves; (b) ECE interviews from the preschool data collection waves; (c) CCOs from the preschool data

collection wave. Standard errors for the present study were estimated using the Taylor Series linearization method (Rust, 1985) in either SAS™ 9.3 or Mplus 7 (Muthén & Muthén, 1998-2012).

Analyses

The primary purpose of the present study was to examine measurement invariance of the ECERS-R (Harms et al., 1998) and CIS (Arnett, 1989) across INC and NSN classrooms and to determine the extent to which there were differences in latent variable factor scores and composite scores for each measure across these two types of classrooms. To examine measurement invariance, MG-CFAs (Jöreskog, 1971) were conducted for each instrument separately. Different analytic strategies were used to examine differences in latent variable factor scores and derived composite scores across INC classrooms. A description of the MG-CFAs and the analyses conducted to examine group differences in latent variables for each instrument is provided.

Multiple group confirmatory factor analyses

For each instrument of interest in the present study, a confirmatory measurement model was proposed based on a review of the literature. The proposed measurement model for the ECERS-R (Harms et al., 1998) was a three-factor model (see Figure A-1) derived primarily from a previous study in which exploratory factors analyses were conducted using ECERS-R data from a similar analytic sample as the present study (Gordon et al., 2013). The proposed measurement model for the CIS (Arnett, 1989) was a bi-factor model (see Figure A-2) derived primarily from a previous study in which exploratory factor analyses were conducted using CIS data from a sub-sample of center-based ECE settings observed as part of the ECLS-B (Colwell et al., 2012). Configural invariance models were tested to examine the fit of the structural models for

the ECERS-R and the CIS. Two MG-CFAs each were conducted separately for the ECERS-R and the CIS. A description of these two models is provided below.

Configural invariance model

For each instrument, a configural invariance model (Horn et al., 1983) was tested to examine whether the instrument measured of the same number of latent variables in INC and NSN classrooms (Research Question 1) and whether the pattern of associations between items and latent variables was equivalent across INC and NSN classrooms (Research Question 2). To test for configural invariance, the pattern of zero and non-zero factor loadings was constrained to be equal across the two types of classrooms. The information gained from testing the configural invariance model was used to determine whether changes the proposed measurement model for the instrument were necessary. In the present study, model fit indices indicated good fit of the measurement models shown in Figures A-1 and A-2, therefore, no changes were made to either measurement model.

Strict invariance

Following confirmation of configural invariance, strict invariance models, in which factor loadings, item thresholds, and residual variances were constrained to be equal across INC and NSN classrooms (Research Question 3), were tested for the ECERS-R (Harms et al., 1998) and CIS (Arnett, 1989). A decision was made in the present study to begin with the most restricted invariance model and release factor loading and item threshold parameters as necessary to establish adequate model fit. Although it is common to test for equivalence of factor loadings (i.e., weak invariance), item thresholds (i.e., strong invariance), and residual variances (i.e., strict invariance) in a stepwise fashion by examining changes in model fit as more restrictions are imposed on

the model (Vandenberg & Lance, 2000; Steenkamp & Baumgartner, 1998), the approach used in the present study has been cited as a viable approach for examining measurement invariance for ordered-categorical measures (Millsap & Yun-Tein, 2004). The strict invariance model was chosen for the present analyses because the restrictions imposed in the model are on the key parameters of interest for examining measurement invariance for ordered-categorical measures (Millsap & Yun-Tein, 2004).

Group Comparisons of Latent Variables

The validity of conducting group comparisons of latent variables on the ECERS-R (1998) and the CIS (1989) was contingent on evidence of measurement invariance for these instruments. Because such evidence was established, analyses were conducted to determine whether there were differences in latent variables and derived composite scores for either instrument across INC and NSN classrooms (Research Question 4). To examine differences in latent variables, the strict invariance model was specified such that the factor means and variances for NSN classrooms were set equal to 0 and 1, respectively, and the factor means and variances for INC classrooms were left free. The described model specifications allowed for estimation of standardized mean differences in factor scores for INC classrooms compared to NSN classrooms.

Additional analyses were conducted to examine differences in derived composite scores. For the present study, the item sets associated with each of the latent variables shown in the measurement models for the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989) are referred to as latent variable subscales. Derived composite scores were calculated by averaging the item-level scores for each latent variable subscale. Three regression analyses were conducted each for the ECERS-R and the CIS. For each regression analysis, group membership was entered as a predictor variable of the

derived composite score for one of the latent variable subscales. Univariate regression analyses were conducted using the Proc SURVEYREG procedures in SAS™ 9.3.

Summary

The present study involved secondary analyses of a subsample of cross-sectional data from the ECLS-B. The primary purpose of the present study was to examine measurement invariance for the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989) across INC and NSN classrooms. A secondary goal of the present study, which was contingent on evidence supporting measurement invariance, was to examine the extent to which there were differences in empirically validated latent variable scores for each instrument across INC and NSN classrooms. MG-CFAs were conducted separately for each instrument to examine measurement invariance across INC and NSN classrooms (Research Questions 1 through 3) and to determine whether there were group differences in latent variables (Research Question 4). Univariate regression analyses were conducted to determine whether there were differences in derived composite scores for empirically validated latent variable subscales across INC and NSN classrooms (Research Question 4).

Table 3-1. Internal Consistency Reliability Scores for Published ECERS-R and CIS Subscales

Subscale	No. Items	Total Sample (N = 1350)	INC classrooms (n = 900)	NSN (n = 450)
ECERS-R				
Total Score	34	.94	.94	.95
Space and Furnishings	8	.78	.75	.80
Personal Care Routines	5	.72	.71	.73
Language-Reasoning	4	.80	.81	.79
Activities	9	.90	.89	.90
Interaction	5	.85	.86	.82
Program Structure	3	.76	.75	.76
CIS				
Total Score	26	.95	.95	.94
Sensitivity	10	.94	.94	.93
Harshness	9	.87	.88	.86
Detachment	4	.77	.78	.75
Permissiveness	3	.59	.61	.54

Note. Estimates weighted by W33P0; only cases with a valid weight were included in analyses

Table 3-2. Weighted Percentage of ECLS-B Focal Children in INC and NSN classrooms

Variable	INC (n = 900)	NSN (n = 450)
Male	50.4	49.5
Ethnicity		
White	54.4	53.7
Black	15.8	17.3
Hispanic	22.3	23.4
Asian	3.1	3.2
Pacific Islander/American Indian/Alaska Native	0.6	0.5
Multi-racial	3.8	2.0
Below poverty threshold	28.6	19.9
SES		
1 st Quintile (lowest)	20.1	15.1
2 nd Quintile	20.2	17.7
3 rd Quintile	19.9	17.2
4 th Quintile	20.9	19.2
5 th Quintile	18.9	30.8
Special needs	40.7	0
Very Low Birth Weight (<1500 g)	1.9	0
EI/ECSE Services	8.8	0
Medical Diagnosis	22.9	0
Professional Evaluation	27	0

Note. Estimates weighted by W33P0; only cases with a valid weight were included in analyses.

^aIncludes children who met any of the four criteria for being characterized with special needs; some children met multiple criteria, therefore, the percentages from types of special needs identified sum to a total greater than the total percentage of children identified with special needs.

Table 3-3. Percentage of Center-Based Care Types by Classroom Group Classification

Care Type	INC (n = 900)	NSN (n = 450)
Public prekindergarten	22	17.2
Private prekindergarten	19	23.8
Child care center	9.6	17.4
Head Start	24.3	8.1
Preschool/nursery school	23.6	31
Other	1.7	2.6

Note. Estimates weighted by W33P0; only cases with a valid weight were included in analyses.

Table 3-4. Mean Scores on Published ECERS-R and CIS Subscales

Published Subscale	Total Sample (N = 1350)	INC (n = 900)	NSN (n = 450)
ECERS-R			
Total Score	4.54 (1.08)	4.66 (1.06)	4.33 (1.09)
Space and Furnishings	4.70 (1.12)	4.86 (1.05)	4.44 (1.19)
Personal Care Routines	3.96 (1.52)	4.12 (1.53)	3.70 (1.45)
Language-Reasoning	4.96 (1.38)	5.07 (1.38)	4.78 (1.34)
Activities	3.91 (1.22)	4.04 (1.16)	3.70 (1.29)
Interaction	5.53 (1.39)	5.56 (1.44)	5.48 (1.30)
Program Structure	4.96 (1.58)	5.08 (1.54)	4.77 (1.64)
CIS			
Total Score	64.57 (11.71)	64.73 (12.07)	64.35 (11.13)
Sensitivity	22.09 (6.32)	22.35 (6.35)	21.70 (6.24)
Harshness	24.03 (3.77)	23.95 (3.98)	24.16 (3.43)
Detachment	10.94 (1.84)	10.64 (1.93)	10.95 (1.70)
Permissiveness	7.51 (1.54)	7.49 (1.61)	7.54 (1.41)

Note. Estimates weighted by W33P0; only cases with a valid weight were included in analyses; *SDs* were generated using the Proc SURVEREG procedure in SAS™ 9.3 and reflect the root mean standard error.

CHAPTER 4 RESULTS

In the present study, secondary correlational analyses were conducted with cross-sectional data from the ECLS-B to examine measurement invariance of the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989) and to determine whether there were differences in latent variables for each instrument across INC and NSN classrooms. Multiple group confirmatory factor analyses were conducted in *Mplus* 7 (Muthén & Muthén, 1998-2012) to examine measurement invariance for each instrument and to determine whether there were differences in the mean factor scores across INC and NSN classrooms. Regression analyses were conducted to examine group differences in derived composite scores using the Proc SURVEYREG procedure in SAS™ 9.3. The present chapter begins with a description of the context for reporting and interpreting findings. For each research question addressed in the present study, information is provided about the analytic procedures used to conduct analyses and interpret findings, followed by a presentation of findings.

Context for Reporting and Interpreting Findings

ECLS-B data were collected using a clustered, list-frame sampling design to obtain a nationally representative sample of children born in 2001. Data for the present study represent a subsample of ECLS-B participants for which CCOs were conducted at the 4-year data collection wave. Sampling weights were generated by ECLS-B researchers to take into account unequal sampling probabilities for the ECLS-B study and the CCO observation in addition to non-response rates within and across data collection waves. All findings reported were estimated using Taylor Series linearization to account for the clustered sampling design (Rust, 1985). Although the data for the

present study were drawn from a nationally representative sample of children, the findings presented are not nationally representative of INC and NSN classrooms; rather, findings from the present study represent classrooms in which a nationally representative sample of 4-year children who received center-based care in either INC or NSN classrooms for more than 10 hours per week were enrolled. To meet IES reporting requirements for restricted-use data sets, all sample sizes have been rounded to the nearest 50.

Research Questions 1 and 2

The purpose of Research Questions 1 and 2 was to determine (a) the extent to which each instrument of interest in the present study measured the same number of latent variables across INC and NSN classrooms and (b) the extent to which the items defining these variables were equivalent across INC and NSN classrooms. For the present study, a three-factor model was proposed for the ECERS-R, which included factors related to the provisions of activities and materials, language and interactions in the classroom, and personal care and safety routines (see Figure A-1 and Table A-1 for a visual depiction of the model and a description of the items associated with each factor). A bi-factor model was proposed for the CIS (see Figure A-2). The CIS model consisted of a general factor onto which every item loaded (i.e., Caregiver Interactions) and two methodological factors that were derived based on the positive or negative orientation of items. The proposed confirmatory models were based on a review of empirical literature regarding factor structures of the ECERS-R and the CIS. Models consistent with the published subscale structure for each instrument were not tested due to a lack of empirical evidence supporting these structures. The confirmatory models proposed for the present study were tested using multiple group confirmatory

factor analyses in *Mplus* 7 (Muthén & Muthén, 1998-2012). To address Research Questions 1 and 2, a configural invariance model was tested for each instrument. Testing configural invariance models provided information about (a) whether the proposed measurement models fit the data INC and NSN classrooms accurately and (b) whether the number of latent variables and pattern of item loadings to latent variables was equivalent across INC and NSN classrooms. The procedures used to estimate the configural invariance models and to evaluate model fit are provided below, followed by research findings regarding model fit.

Estimation of Configural Invariance Models

Item responses on the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989) were specified to be on ordered categorical scales and the configural invariance models for the these instruments were estimated using diagonally Weighted Least Squares estimation in *Mplus* 7 (Muthén & Muthén, 1998-2012). The theta parameterization was used. In each configural invariance model, the number of latent variables and the pattern of items defining the latent variables were specified to be the same across groups. The specification of the model used the default approach described in Muthén and Muthén (1998-2012): factor loadings that were not specified to be zero and thresholds were not constrained to be equal across groups; residual variances were specified to be equal to one in both groups; factor means were specified to be equal in the two groups; unit factor variances were specified for NSN classrooms and factor variances were freely estimated for INC. By specifying the residual variances to equal one, the model is equivalent to the latent variable specification of a graded response model (GRM) (see Muthén and Asparouhov, 2002). The GRM is an item response theory model for item scores recorded on ordered categorical scales.

For the CIS model, the two methodological factors were specified to be uncorrelated with the Caregiver Interactions factor. The syntax used to estimate the configural invariance models is provided in Appendix D.

Evaluating Model Fit for Configural Invariance Models

Model fit indices were used to examine model fit for the ECERS-R (Harms et al., 1998) and CIS (Arnett, 1989) configural invariance models. The indicator of absolute model fit that was used in the present study was χ^2 . The recommended criterion for establishing adequate model fit using χ^2 is an insignificant result at the .05 threshold (Barret, 2007). Although the *Chi-square* test is a popular criterion for examining model fit, it was not used as the only criterion for evaluating model fit in the present study, because there is a tendency for this test to result in rejection of well-fittingmodels when sample sizes are large (Bentler & Bonnet, 1980; Jöreskog & Sorbom, 1993). In addition to using the *Chi-square* test, the criteria described by Hu and Bentler (1999) for the root mean square error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker-Lewis index (TLI) were used. RMSEA estimates provide information about model fit with respect to the population covariance matrix (Byrne, 1998). CFI and TLI estimates provide information about the fit of the model based on a comparison of the χ^2 statistics for the proposed model and a null model in which all items are hypothesized to be uncorrelated (Bentler, 1990; Bentler & Bonnet, 1980; Byrne, 1998). The Hu-Bentler criteria suggest RMSEA values $\leq .06$, and CFI and TLI values $\geq .95$ to indicate good model fit (Hu & Bentler, 1999).

ECERS-R Model Fit

Model fit indices for the ECERS-R (Harms et al., 1998) configural invariance model are shown in Table 4-1. The *Chi-square* test for this model indicated inadequate fit of the model to the data. Despite a significant *Chi-square* test, RMSEA, CFI, and TLI estimates met the recommended criteria proposed by Hu and Bentler (1999). Given the sample size for the present study, it is likely the incongruence between the *Chi-square* test and the comparative fit indices is due to sample size. As such, the RMSEA, CFI, and TLI indices were used as the primary source of information to determine the adequacy of the configural invariance model. The RMSEA for the ECERS-R configural invariance model was well within the limits recommended by Hu and Bentler (1999) at .029 with a 9% confidence interval of .027 to .032. The CFI (.946) and the TLI (.942) estimates were slightly below the criteria proposed by Hu and Bentler (1999) but still suggested adequate model fit for the ECERS-R configural invariance model. Given this evidence, it was determined that no changes to the proposed three-factor model were necessary.

CIS Model Fit

Model fit indices for the ECERS-R (Harms et al., 1998) configural invariance model are shown in Table 4-1. The *Chi-square* test for the CIS (Arnett, 1989) configural invariance model was significant, indicating inadequate fit of the model to the data. RMSEA, CFI, and TLI estimates indicated good model fit (RMSEA = .029, CFI = .984, TLI = .981). Given that RMSEA, CFI, and TLI were well within the limits proposed by Hu and Bentler (1999), it was determined that no changes to the proposed bi-factor model were necessary.

Research Question 3

The purpose of Research Question 3 was to determine the extent to which factor loadings, item thresholds, and residual variances for the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989) were equivalent across INC and NSN classrooms.

Equivalence of these parameters across INC and NSN classrooms suggests that in addition to the factor structure being equivalent across groups (i.e., configural invariance), the metric on which the factors are measured is also equivalent. Such evidence is important because it suggests latent variables are comparable across groups; thus, findings from comparisons of the latent variables can be interpreted meaningfully to make inferences about potential differences on the latent variables identified in the model. To address Research Question 3, strict invariance models were fit separately for each instrument using *Mplus* 7 (Muthén & Muthén, 1998-2012).

Because the strict invariance model provides the strongest evidence of measurement invariance, a decision was made to begin with this model and release restrictions on pairs of factor loadings and item thresholds as necessary to obtain good model fit rather than examining changes in model fit for the configural, weak, strong, and strict invariance models. The procedure used in the present study has been described as a viable procedure for examining measurement invariance for ordinal-categorical measures such as the ECERS-R and the CIS (Millsap & Yun-Tein, 2004). The procedures used to estimate the strict invariance models and to evaluate model fit are described below, followed by a presentation of findings.

Estimation of Strict Invariance Models

The strict invariance models for the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989) were estimated using diagonally Weighted Least Squares estimation in

Mplus 7 (Muthén & Muthén, 1998-2012). As in the configural invariance models, the number of latent variables and pattern of items associated with latent variables across INC and NSN classrooms was the same for both groups and the theta parameterization was used. In addition, factor loadings and items thresholds were constrained to be equal across the two groups and residual variances were set equal to one in both groups. As a result of the latter specification, the estimate models were equivalent to invariant GRM models. The impact of the setting the factor means equal to zero in one group and allowing them to be estimated in the second group allowed for a comparison of latent variable means across groups. The syntax used to estimate the strict invariance models is provided in Appendix D.

Evaluating Model Fit for the Strict Invariance Models

To evaluate model fit for the strict invariance models, χ^2 , RMSEA, CFI, and TLI model fit indices were examined using the same criteria applied for the configural invariance models. The statistical significance of non-standardized factor loadings and the magnitude of standardized factor loadings for each item were examined to determine the strength of association between items and latent variables. Standardized factor loadings greater than .40 were considered indicative of a strong association between the item and the latent variable. In addition, the proportion of classrooms receiving each of the possible item scores for the instruments were examined in INC and NSN classrooms to determine whether the pattern of proportions for item scores were similar across groups. Item score proportions are directly related to item thresholds, such that each item threshold for each item represents a cut point on a normal curve and the proportion of classrooms earning each of the item score

categories represents the area under the curve between item thresholds (Algina, personal communication, June 27, 2013). For the strict invariance model, item score proportions are related to differences between the groups in the factor means. Unless the means are dramatically different, the item score proportions should be similar for the two groups. Therefore, given the relationship between item score proportions and item thresholds, similarity in the proportions of item scores across INC and NSN classrooms provides evidence of equivalent item thresholds across groups. In addition, the item score proportions provide evidence about how the groups differ on item scores.

ECERS-R Model Fit

The *Chi-square* test for the ECERS-R (Harms et al., 1998) strict invariance model indicated inadequate fit of the model to the data. The RMSEA, CFI, and TLI indices of model fit met the recommended criteria proposed by Hu and Bentler (1999; see Table 4-1 for estimates of model fit indices). The RMSEA for the ECERS-R strict invariance model was well within the limits recommended by Hu and Bentler (1999) at .024 with a 95% confidence interval of .022 to .027. The CFI (.988) and the TLI (.989) also indicated good model fit for the strict invariance ECERS-R Model. Given the evidence for good model fit based on these indices, it was determined that it was not necessary to release any parameters in the model.

Non-standardized factor loadings for the ECERS-R strict invariance model are shown in Table 4-2. Because the strict invariance model constrained factor loadings to be equivalent across INC and NSN classrooms, the factor loadings presented in Table 4-2 represent factor loadings for both groups. All of the factor loadings were statistically significant. In addition to statistically significant factor loadings, a review of standardized

factor loading revealed strong associations between items and latent variables. Table 4-3 displays the standardized factor loadings for INC classrooms. Although standardized factor loadings differed slightly across INC and NSN classroom due to factor variances, these differences were minimal, and the factor loading presented in Table 4-3 are representative of standardized factors loadings for both groups of classrooms. As shown in Table 4-3, standardized factor loadings for the Activities and Materials factor ranged from 0.469 to .788. Standardized factor loadings for the Language and Interactions factor ranged from .566 to .910, with all except one loading above .70. Standardized factor loadings for the Personal Care and Safety factor ranged from .580 to .728. The Activities and Material factor was correlated fairly highly with the Language and Interactions factor and with the and the Personal Care and Safety factor, with correlations of .76 and .64, respectively. The Language and Interactions and Personal Care and Safety factors were also correlated with each other (.61). Internal consistency score reliabilities for the item sets associated with each of these three factors ranged from .77 (Personal Care and Safety) to .93 (Activities and Materials) in the total analytical sample (see Table 4-4). As shown in Table 4-4, internal consistency score reliabilities for each of the three factors were similar across INC (*range* = .76 - .92) and NSN classrooms (*range* = .77 - .94).

The proportions of classrooms receiving each of the possible item scores for the ECERS-R are presented in Tables 4-5 (i.e., INC) and 4-6 (i.e., NSN). Although there was some variability in the exact proportions, the pattern of proportions was similar across the two types of classrooms. In general, the score categories with the highest proportions for items associated with the Activities and Materials factor were categories

3 and 4. The highest proportions of scores for items associated with the Language and Interactions factor were generally either in the 3-4 range or in the 6-7 range. Scores for items associated with the Personal Care and Safety factor were generally the lowest, with high proportions of classrooms receiving scores in the 1-2 range. Given the similarities in the proportions of item scores across INC and NSN classrooms, the assertion of equivalent item thresholds across the groups, which was based on model fit indices, is reasonable.

CIS Model Fit

Model fit indices for the CIS strict invariance model are provided in Table 4-1. The *Chi-square* test for the CIS (Arnett, 1989) strict invariance model was significant, indicating inadequate fit of the model to the data; however, RMSEA, CFI, and TLI estimates indicated good model fit (RMSEA = .023, CFI = .988, TLI = .989). Given that RMSE, CFI, and TLI were well within the limits proposed by Hu and Bentler (1999), it was determined that it was not necessary to release any parameters in the model.

The non-standardized factor loadings for the CIS strict invariance model are presented in Table 4-7. All of the factor loadings for the Caregiver Interactions factor were statistically significant. With the exception of four items (i.e., seems distant or detached, punishes without explanation, not interested in children's activities, doesn't supervise closely), factor loadings for the Negative Orientation factor were statistically significant. All of the factor loadings for the Positive Orientation factor were statistically significant. Standardized factor loadings for the INC classrooms are presented in Table 4-8. Although standardized factor loadings differed slightly across INC and NSN classroom due to factor variances, these differences were minimal, and the factor loadings presented in Table 4-8 are representative of standardized factors loadings for

both groups of classrooms. Standardized factor loadings for the Caregiver Interactions Factor ranged from .310 to .888. With the exception of one item (i.e., expects self-control), the standardized loadings for all of the items on this factor were greater than the criterion of .40, indicating a strong association with the latent variable. Standardized factor loadings for the Negative Orientation factor ranged from -.319 to .787. The associations between items and the latent variable were not as strong for this factor, with several items having standardized loadings below .40. A similar pattern was observed for the Positive Orientation factor. Standardized factor loadings ranged from .189 to .666, with only four items with standardized loadings above .40 (i.e., explains rules, talks to children on a level they understand, firm when necessary, encourages pro-social behavior). Internal consistency score reliabilities for the three factors ranged from .91 to .95 for the total sample of classrooms and were nearly equivalent across INC and NSN classrooms (see Table 4-9).

The proportions of classrooms receiving each of the item scores for the CIS are presented in Tables 4-10 (i.e., INC) and 4-11 (i.e., NSN). There was limited variability in item-level scores within and across INC and NSN classrooms. The highest proportion of scores received in the 3-4 range for all items across INC and NSN classrooms. Similarities in the proportions of item scores across INC and NSN classrooms substantiate the claim that the item thresholds are equivalent across groups.

Research Question 4

The purpose of this question was to examine group differences in the empirically derived latent variables for the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989). After establishing evidence of measurement invariance for each instrument, analyses were conducted to examine group differences in latent variable factor score

means and group differences in derived composite score means. The procedures used to examine group differences in these two variables are reported below, followed by a presentation of research findings.

Examining Group Differences in Latent Variable Scores

Two different approaches were used to examine group differences in latent variable scores for the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989). First, mean differences in factor scores were examined based on findings from the estimation of the strict invariance model for each instrument. In the strict invariance model, factor means for NSN classrooms were set equal to zero and factor variances were set equal to 1. Factor factor means and variances for INC were left free, which allowed for a comparison of mean factor scores across groups. Because the factor means for NSN classrooms were set equal to zero and the factor variances were set equal to 1, the estimate of the means for INC were non-pooled *Cohen's d* estimates, which were calculated using the standard deviation of the NSN classrooms as the divisor.

The second approach used to examine mean differences in latent variables across groups was to conduct regression analyses in which group membership was the predictor and continuous derived composite scores were the outcomes. Derived composite scores were computed by averaging the item-level scores for the items associated with each of the factors identified in the measurement models for the ECERS-R (Harms et al., 1998) and CIS (Arnett, 1989). In the present discussion, derived composite scores are also referred to as latent variable subscale scores to distinguish them from the subscales published by the instrument developers.

Regression analyses were conducted using the Proc SURVEYREG procedure in SAS™ 9.3. A separate regression analysis was conducted for each of the factors. Effect sizes

were calculated by subtracting the group mean for NSN classrooms from the group mean for INC classrooms and dividing the result by the NSN classroom standard deviation. Findings from these analyses are presented in this section. Findings are discussed with respect to each instrument separately.

Mean differences in ECERS-R latent variables

There were statistically significant differences in factor scores for all three ECERS-R (Harms et al., 1998) latent variables. On average, factor scores for INC classrooms were higher than those for NSN classrooms. Effect sizes for the Activities and Materials, Language and Interactions, and Personal Care and Safety factors were 0.27 ($p = .015$), 0.29 ($p = .049$), and .29 ($p = .022$), respectively. Table 4-12 displays the group means and effect sizes for each of the latent variable subscales. Mean scores for derived composites were higher in INC classrooms for all three latent variable subscales for the ECERS-R. There were statistically significant differences for the Activities and Materials [$F(53) = 6.23, p = .016$] and Personal Care and Safety [$F(53) = 5.67, p = .029$] composite scores. Differences in the Language and Interactions derived composite scores were not statistically significant [$F(53) = 2.34, p = .132$]. The effect sizes for mean differences in derived composite scores were similar to those for the comparison of latent variable factor scores and were in the small to moderate range (see Table 4-10).

Mean differences in CIS latent variables

There were no statistically significant differences in either the factor scores or the derived composite scores for the CIS. Effect sizes for factor scores were very small. The effect size for the caregiver interactions factor was 0.09; the effect size for the factor associated with negatively oriented items was 0.04; and the effect size for the

factor associated with positively oriented items was -.0.09, indicating almost no difference in factor scores across groups. Table 4-12 displays the group means and effect sizes for each of the latent variable subscales for the CIS. Mean scores for derived composites were nearly equivalent for all three latent variable subscales across groups. The effect sizes for mean differences in derived composite scores were similar to those for the comparison of latent variable factor scores, indicating no differences statistically significant or notable differences on CIS latent variables.

Summary

Results from the present study provide evidence to support the hypothesis that the ECERS-R measures three latent variables associated with quality in ECE (i.e., Activities and Materials, Language and Interactions, Personal Care and Safety). Findings from the multiple group factor analyses provided evidence to suggest that these three variables are measured equally across INC and NSN classrooms, permitting comparisons of latent variables across these two types of classrooms. Comparisons of latent variable scores revealed small to moderate differences in the factor scores and derived composite scores for all three latent variables across groups. Group differences in all latent variable factor scores were statistically significant. Group difference in the derived composite scores for Activities and Materials and for Personal Care and Safety were also statistically significant, indicating higher quality in INC classrooms.

Study findings related to the CIS (Arnett, 1989) provided evidence to suggest CIS provides a global measure of caregiver interactions, when controlling for two methodological factors associated with the negative or positive orientation of item wording. There was evidence to suggest that measurement of both the substantive

(i.e., Caregiver Interactions) and the methodological factors was equivalent across INC and NSN classrooms, permitting comparisons of latent variable scores across groups. Comparisons of latent variable factor scores and derived composite scores did not reveal statistically significant or notable differences in latent variable scores across INC and NSN classrooms.

The present study offers preliminary evidence to suggest both the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989) provide equivalent measures of quality dimensions across INC and NSN classrooms. Findings from the present study also add to the literature base concerning differences in global, structural, and process quality in classrooms with different compositions of children with special needs.

Table 4-1. Model Fit Statistics for Multiple Group Confirmatory Factor Models

Model	χ^2 (df)	RMSE (90% CI)	CFI	TLI
ECERS-R				
Configural Invariance	1652.29* (1048)	.029 (.027 - .032)	.946	.942
Strict Invariance	1786.33* (1280)	.024 (.022 - .027)	.955	.960
CIS				
Configural Invariance	849.91* (544)	.029 (.025 - .033)	.984	.981
Strict Invariance	899.76* (668)	.023 (.019 - .026)	.988	.989

Note. Estimates weighted by WPP30. Only cases with valid weights were included in analyses; * $p < .05$

Table 4-2. Non-Standardized Factor Loadings from the ECERS-R Strict Invariance Model.

Item	Activities and Materials	Language and Interactions	Personal Care and Safety
1. Indoor space	0.59*	--	--
2. Furnishings routine care/learning	0.74*	--	--
3. Furnishings relaxation	0.98*	--	--
4. Room arrangement	1.15*	--	--
5. Space for privacy	0.96*	--	--
6. Child display	0.79*	--	--
8. Gross motor equipment	0.54*	--	--
15. Books/pictures	1.07*	--	--
19. Fine motor	1.41*	--	--
20. Art	1.34*	--	--
21. Music/movement	0.95*	--	--
22. Blocks	1.32*	--	--
23. Sand/water	0.94*	--	--
24. Dramatic play	1.09*	--	--
25. Nature/science	1.39*	--	--
26. Math/numbers	1.28*	--	--
28. Promoting diversity	0.73*	--	--
34. Schedule	0.87*	--	--
35. Free play	1.35*	--	--
36. Group time	1.25*	--	--
9. Greeting/departure	--	0.53*	--
16. Encouraging communication	--	1.70*	--
17. Language-reasoning	--	0.93*	--
18. Informal language	--	1.26*	--
29. Gross motor supervision	--	0.77*	--
30. General supervision	--	0.95*	--
31. Discipline	--	1.33*	--
32. Staff-child interactions	--	1.25*	--
33. Child interactions	--	1.24*	--
7. Gross motor space	--	--	0.77*
10. Meals/snacks	--	--	1.15*
12. Toileting/diapering	--	--	1.05*
13. Health practices	--	--	0.97*
14. Safety practices	--	--	1.09*

Note. Estimates weighted by WPP30. Only cases with a valid weight were included in analysis; * $p < .05$

Table 4-3. Standardized Factor Loadings for INC Classrooms from the ECERS-R Strict Invariance Model

Item	Activities and Materials	Language and Interaction	Personal Care and Safety
1. Indoor space	.469*	--	--
2. Furnishings routine care/learning	.555*	--	--
3. Furnishings relaxation	.663*	--	--
4. Room arrangement	.721*	--	--
5. Space for privacy	.656*	--	--
6. Child display	.580*	--	--
8. Gross motor equipment	.436*	--	--
15. Books/pictures	.696*	--	--
19. Fine motor	.788*	--	--
20. Art	.770*	--	--
21. Music/movement	.650*	--	--
22. Blocks	.767*	--	--
23. Sand/water	.647*	--	--
24. Dramatic play	.702*	--	--
25. Nature/science	.783*	--	--
26. Math/numbers	.756*	--	--
28. Promoting diversity	.552*	--	--
34. Schedule	.618*	--	--
35. Free play	.773*	--	--
36. Group time	.749*	--	--
9. Greeting/departure	--	.566*	--
16. Encouraging communication	--	.910*	--
17. Language-reasoning	--	.766*	--
18. Informal language	--	.850*	--
29. Gross motor supervision	--	.705*	--
30. General supervision	--	.775*	--
31. Discipline	--	.862*	--
32. Staff-child interactions	--	.849*	--
33. Child interactions	--	.846*	--
7. Gross motor space	--	--	.580*
10. Meals/snacks	--	--	.728*
12. Toileting/diapering	--	--	.694*
13. Health practices	--	--	.668*
14. Safety practices	--	--	.710*

Note. Estimates weighted by WPP30. Only cases with a valid weight were included in analysis; * $p < .05$

Table 4-4. Internal Consistency Reliabilities for ECERS-R Latent Variable Subscales

Latent Variable Subscale	No. Items	Total Sample	INC	NSN
Activities and Materials	20	.93	.92	.94
Language and Interactions	9	.89	.90	.87
Personal Care and Safety	5	.77	.76	.77

Note. Estimates weighted by W33P0; only cases with a valid weight were included in analyses.

Table 4-5. Percentage of Scores for Each ECERS-R Score Category in NSN Classrooms.

Item	1	2	3	4	5	6	7
Activities and Materials							
1. Indoor space	4.8	6.3	8.8	28.5	7.3	15.4	28.9
2. Furnishings routine care/learning	4.3	1.8	0.8	10.8	1.5	27.1	53.6
3. Furnishings relaxation	9.0	4.3	29.4	25.5	5.0	10.1	16.9
4. Room arrangement	7.9	10.9	10.1	15.1	9.8	14.5	31.7
5. Space for privacy	11.5	4.9	32.2	19.8	7.4	8.4	15.9
6. Child display	4.6	12.7	24.1	25.5	8.4	16.5	8.2
8. Gross motor equipment	12.6	19.6	8.2	19.9	3.7	14.1	22.0
15. Books/pictures	2.6	7.7	8.9	57.8	0.9	2.5	19.4
19. Fine motor	5.8	7.8	13.8	34.5	2.7	12.3	23.0
20. Art	5.6	15.0	22.0	34.3	3.4	9.0	10.7
21. Music/movement	2.1	32.8	14.3	31.4	8.1	6.9	4.4
22. Blocks	17.2	8.0	5.1	39.8	7.7	19.6	2.6
23. Sand/water	29.4	3.0	16.4	22.6	4.7	16.8	7.1
24. Dramatic play	10.9	12.7	14.7	39.0	10.8	9.5	2.4
25. Nature/science	25.6	19.5	6.8	34.0	0.2	5.0	8.9
26. Math/numbers	12.4	2.7	18.7	44.5	3.0	6.6	12.0
28. Promoting diversity	5.4	13.1	27.4	30.0	6.3	7.8	10.0
34. Schedule	2.9	34.4	4.2	25.5	1.5	5.2	26.3
35. Free play	7.6	7.4	7.7	24.4	5.3	16.3	31.4
36. Group time	10.1	1.1	5.7	14.5	4.9	18.0	45.7
Language and Interactions							
9. Greeting/departure	3.2	6.0	3.2	14.0	2.5	13.5	57.5
16. Encouraging communication	3.3	2.6	11.4	22.5	3.6	19.7	37.1
17. Language-reasoning	5.9	11.5	22.4	22.4	6.1	7.9	23.8
18. Informal language	2.9	1.9	9.7	30.7	2.8	12.6	39.4
29. Gross motor supervision	6.8	8.1	2.9	21.8	22.4	22.8	15.2
30. General supervision	4.6	5.6	3.6	12.1	12.9	19.7	41.6
31. Discipline	3.8	9.7	2.4	11.7	17.5	26.5	28.4
32. Staff-child interactions	3.5	3.7	1.2	10.5	0.3	10.3	70.3
33. Child interactions	2.6	4.7	1.7	15.1	1.6	25.1	49.3
Personal Care and Safety							
7. Gross motor space	11.6	26.3	8.4	21.7	9.3	16.4	6.3
10. Meals/snacks	43.3	19.4	0.9	13.7	5.6	6.5	10.6
12. Toileting/diapering	30.5	25.5	0.9	13.8	5.6	6.5	10.6
13. Health practices	8.9	55.7	0.3	7.8	1.2	12.1	14.1
14. Safety practices	31.3	18.2	0.8	12.7	1.3	6.2	29.4

Note. Estimates weighted by WPP30. Only cases with a valid weight were included in analysis; higher scores indicate higher quality.

Table 4-6. Percentage of Scores for Each ECERS-R Score Category in INC Classrooms.

Item	1	2	3	4	5	6	7
Activities and Materials							
1. Indoor space	2.4	4.4	4.5	27.1	4.3	12.5	44.9
2. Furnishings routine care/learning	2.9	1.2	0.0	6.3	4.8	19.7	65.1
3. Furnishings relaxation	6.6	3.9	13.9	39.6	7.7	10.8	17.4
4. Room arrangement	3.4	4.8	9.5	12.8	5.3	13.9	50.3
5. Space for privacy	5.8	5.1	20.3	32.5	5.7	11.7	18.9
6. Child display	1.6	7.4	22.7	32.5	9.5	14.1	12.3
8. Gross motor equipment	9.0	19.2	3.0	17.6	6.0	16.1	29.2
15. Books/pictures	3.1	4.6	7.0	59.3	2.1	3.0	20.8
19. Fine motor	2.4	5.2	7.0	47.2	3.9	8.9	25.4
20. Art	4.8	7.0	20.5	39.8	4.6	7.5	15.8
21. Music/movement	5.8	27.3	14.5	30.9	5.8	6.6	9.2
22. Blocks	6.0	8.2	5.0	51.3	7.0	16.8	5.7
23. Sand/water	12.9	4.7	17.0	30.2	8.3	13.9	12.9
24. Dramatic play	6.6	7.5	11.9	50.4	9.6	11.5	2.6
25. Nature/science	14.5	19.4	8.3	43.1	1.2	3.5	10.0
26. Math/numbers	7.3	2.7	14.6	54.3	3.8	4.0	13.2
28. Promoting diversity	3.8	8.0	20.6	32.6	11.6	7.0	16.4
34. Schedule	4.7	25.5	2.3	26.3	1.6	10.5	29.2
35. Free play	6.4	6.1	5.9	25.9	4.9	11.0	39.8
36. Group time	7.1	1.8	7.0	14.1	4.2	14.6	51.1
Language and Interactions							
9. Greeting/departure	2.5	4.1	4.2	8.1	2.8	10.7	67.6
16. Encouraging communication	2.7	4.1	4.1	16.7	4.8	21.1	46.6
17. Language-reasoning	7.3	4.1	17.4	25.8	4.5	8.6	32.3
18. Informal language	2.8	2.2	7.4	27.0	3.0	11.6	46.0
29. Gross motor	7.6	4.5	5.2	16.8	17.1	20.2	28.5
supervision							
30. General supervision	6.7	7.4	2.2	12.0	10.	12.4	48.5
31. Discipline	4.0	4.6	4.3	13.8	11.6	22.0	39.7
32. Staff-child interactions	4.6	5.8	1.9	9.3	1.3	5.3	71.7
33. Child interactions	4.0	4.0	2.3	11.6	2.3	26.1	49.7
Personal Care and Safety							
7. Gross motor space	9.3	24.9	6.8	20.7	10.9	14.6	12.8
10. Meals/snacks	39.2	17.3	0.2	7.9	5.8	13.1	16.5
12. Toileting/diapering	29.9	23.0	2.0	9.8	0.5	7.8	27.0
13. Health practices	4.0	45.5	3.6	7.7	4.7	9.1	25.3
14. Safety practices	24.6	23.3	0.8	7.4	1.3	7.5	35.3

Note. Estimates weighted by WPP30. Only cases with a valid weight were included in analysis; higher scores indicate higher quality.

Table 4-7. Non-Standardized Factor Loadings from the CIS Strict Invariance Model

Item	Caregiver Interactions	Negative Orientation	Positive Orientation
1. Speaks warmly	2.00*	--	0.51*
2. Seems critical ^a	2.14*	0.99*	--
3. Listens attentively	1.70*	--	0.58*
4. High value on obedience ^a	0.69*	1.10*	--
5. Seems distant or detached ^a	1.68*	0.04	--
6. Enjoys children	1.85*	--	0.43*
7. Explains rules	0.96*	--	0.67*
8. Encourages new experiences	1.23*	--	0.43*
9. Doesn't try to exercise much control ^a	1.79*	2.63*	--
10. Speaks with irritation ^a	1.96*	0.87*	--
11. Enthusiastic about activities and efforts	1.88*	--	0.67*
12. Threatens children ^a	0.97*	0.70*	--
13. Not involved with children ^a	0.83*	-0.43*	--
14. Pays positive attention	1.81*	--	0.73*
15. Doesn't reprimand mishbehavior ^a	1.62*	1.01*	--
16. Talks to children on level they understand	1.34*	--	0.94*
17. Punishes without explanation ^a	1.11*	0.28*	--
18. Firm when necessary	0.71*	--	1.12*
19. Encourages pro-social behavior	1.12*	--	0.78*
20. Finds fault easily ^a	1.93*	1.43*	--
21. Disinterest in children's activities ^a	1.24*	-0.03	--
22. Prohibits many things ^a	1.05*	1.07*	--
23. Doesn't supervise closely ^a	0.85*	-0.15	--
24. Expects self-control	0.41*	--	0.91*
25. Talks to children on their level	0.96*	--	0.48*
26. Harsh when scolding ^a	1.98*	1.34*	--

Note. Estimates weighted by WPP30. Only cases with a valid weight were included in analysis.

^aItems reverse-scored so that a higher score indicates higher quality.

Table 4-8. Standardized Factor Loadings INC Classrooms from the CIS Strict Invariance Model

Item	Factor 1 Caregiver Interactions	Factor 2 Negative Orientation	Factor 3 Positive Orientation
8. Speaks warmly	.888*	--	.208*
9. Seems critical ^a	.848*	.385*	--
10. Listens attentively	.848*	--	.265*
11. High value on obedience ^a	.435*	.684*	--
12. Seems distant or detached ^a	.876*	.022	--
13. Enjoys children	.879*	--	.189*
14. Explains rules	.653*	--	.424*
8. Encourages new experiences	.774*	--	.250*
9. Doesn't try to exercise much control ^a	.549*	.787*	--
10. Speaks with irritation ^a	.843*	.365*	--
11. Enthusiastic about activities and efforts	.860*	--	.286*
12. Threatens children ^a	.647*	.452*	--
13. Not involved with children ^a	.632*	-.319*	--
14. Pays positive attention	.846*	--	.315*
15. Doesn't reprimand mishbehavior ^a	.769*	.466*	--
16. Talks to children on level they understand	.727*	--	.472*
17. Punishes without explanation ^a	.755*	.184*	--
18. Firm when necessary	.453*	--	.666*
19. Encourages pro-social behavior	.691*	--	.447*
20. Finds fault easily ^a	.757*	.545*	--
21. Disinterest in children's activities ^a	.803*	-.021	--
22. Prohibits many things ^a	.601*	.599*	--
23. Doesn't supervise closely ^a	.675*	-.114	--
24. Expects self-control	.310*	--	.639*
25. Talks to children on their level	.684*	--	.314*
26. Harsh when scolding ^a	.772*	.525*	--

Note. Estimates weighted by WPP30. Only cases with a valid weight were included in analysis; * $p < .05$

Table 4-9. Internal Consistency Reliabilities for CIS Latent Variable Subscales

Latent Variable Subscale	No. Items	Total Sample	INC	NSN
Caregiver Interactions	26	.95	.95	.94
Negative Orientation	14	.91	.91	.91
Positive Orientation	12	.93	.93	.93

Note. Estimates weighted by WPP30. Only cases with a valid weight were included in analyses.

Table 4-10. Percentage of Scores for Each CIS Score Category in NSN Classrooms.

Item	1	2	3	4
1. Speaks warmly	1.3	9.3	42.1	47.3
2. Seems critical ^a	1.2	1.8	10.5	86.5
3. Listens attentively	2.6	12.9	42.7	41.8
4. High value on obedience ^a	4.4	11.5	24.7	59.3
5. Seems distant or detached ^a	1.3	2.9	13.6	82.1
6. Enjoys children	1.3	12.0	39.3	47.
7. Explains rules	4.0	31.0	39.1	26.0
8. Encourages new experiences	6.2	21.8	44.2	27.7
9. Doesn't try to exercise much control ^a	2.8	2.5	19.4	75.3
10. Speaks with irritation ^a	1.2	2.8	10.3	85.7
11. Enthusiastic about activities and efforts	4.0	19.6	36.8	39.7
12. Threatens children ^a	1.3	1.1	18.4	79.3
13. Not involved with children ^a	1.2	4.4	16.7	77.7
14. Pays positive attention	2.6	16.4	31.2	49.9
15. Doesn't reprimand mishbehavior ^a	0.9	3.3	15.2	80.6
16. Talks to children on level they understand	2.5	9.4	41.9	46.2
17. Punishes without explanation ^a	1.2	1.8	7.2	89.8
18. Firm when necessary	1.3	19.1	46.7	32.9
19. Encourages pro-social behavior	2.3	14.7	41.0	42.0
20. Finds fault easily ^a	1.3	1.1	8.4	89.2
21. Disinterest in children's activities ^a	1.5	2.0	17.2	79.2
22. Prohibits many things ^a	1.0	4.9	20.2	73.9
23. Doesn't supervise closely ^a	0.4	1.7	22.4	75.6
24. Expects self-control	3.2	12.1	53.6	31.2
25. Talks to children on their level	1.2	24.3	44.9	29.5
26. Harsh when scolding ^a	1.0	1.4	6.3	91.3

Note. Estimates weighted by WPP30. Only cases with a valid weight were included in analysis.

^aItems reverse-scored so that a higher score indicates high quality.

Table 4-11. Percentage of Scores for Each CIS Response Category in INC Classrooms.

Item	1	2	3	4
1. Speaks warmly	0.5	13.6	27.8	58.0
2. Seems critical ^a	0.7	2.0	8.4	89.0
3. Listens attentively	0.9	11.5	43.3	44.4
4. High value on obedience ^a	4.9	8.9	24.2	62.0
5. Seems distant or detached ^a	1.3	2.2	13.9	82.5
6. Enjoys children	2.2	13.6	30.3	53.9
7. Explains rules	3.6	24.8	40.5	31.0
8. Encourages new experiences	4.1	19.7	42.5	33.7
9. Doesn't try to exercise much control ^a	3.9	6.2	16.2	73.7
10. Speaks with irritation ^a	1.4	3.2	12.9	82.4
11. Enthusiastic about activities and efforts	3.4	15.3	37.8	43.5
12. Threatens children ^a	1.7	3.0	10.4	85.0
13. Not involved with children ^a	4.0	1.8	15.0	79.1
14. Pays positive attention	2.1	16.2	30.7	51.0
15. Doesn't reprimand mishbehavior ^a	2.2	2.9	12.1	82.8
16. Talks to children on level they understand	1.4	9.7	39.4	49.6
17. Punishes without explanation ^a	1.9	2.1	9.9	86.0
18. Firm when necessary	3.2	15.7	45.4	35.7
19. Encourages pro-social behavior	3.6	14.5	42.5	39.3
20. Finds fault easily ^a	1.8	2.6	9.1	86.4
21. Disinterest in children's activities ^a	1.8	4.2	13.7	80.3
22. Prohibits many things ^a	3.1	6.6	18.7	71.6
23. Doesn't supervise closely ^a	2.1	2.7	14.3	81.0
24. Expects self-control	5.9	13.6	45.6	35.0
25. Talks to children on their level	2.5	19.3	40.6	37.5
26. Harsh when scolding ^a	1.6	2.6	7.4	88.4

Note. Estimates weighted by WPP30. Only cases with a valid weight were included in analysis.

^aItems reverse-scored so that a higher score indicates higher quality.

Table 4-12. Mean Scores for ECERS-R and CIS Latent Variable Subscales

Latent Variable Subscale	<i>M (SD)</i> Total Sample <i>n</i> = 1350	<i>M (SD)</i> INC <i>n</i> = 900	<i>M (SD)</i> NSN <i>n</i> = 450	<i>F</i>	<i>d</i>
ECERS-R					
Activities and Materials	4.39 (1.11)	4.52 (1.05)	4.19 (1.19)	6.23*	0.28
Language and Interactions	5.45 (1.28)	5.52 (1.21)	5.33 (1.33)	2.34	0.14
Personal Care and Safety	3.60 (1.63)	3.75 (1.66)	3.36 (1.58)	5.67*	0.25
CIS					
Caregiver Interactions	3.48 (0.45)	3.49 (0.46)	3.47 (0.43)	0.08	0.05
Negative Orientation	3.73 (0.41)	3.72 (0.42)	3.74 (0.38)	0.19	-0.05
Positive Orientation	3.19 (0.60)	3.21 (0.60)	3.16 (0.59)	0.58	0.08

Note. Estimates weighted by W33P0; only cases with a valid weight were included in analyses; *SDs* were generated using the Proc SURVEREG procedure in SAS™ 9.3 and reflect the root mean standard error.

CHAPTER 5 DISCUSSION

The primary purpose of the present study was to establish evidence of measurement invariance for the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989) across INC and NSN classrooms. A secondary goal of the present study was, in the event that there was evidence to suggest measurement invariance of the instruments, to determine whether there were differences in latent variables and derived composite scores for each instrument across INC and NSN classrooms. Secondary analyses were conducted with cross-sectional data from the 4-year data collection wave of the ECLS-B. The ECLS-B data set includes data from a nationally representative sample of children born in the U.S. in 2001. The analytic sample for the present study included data from a subsample of children whose INC or NSN classrooms were observed and rated using the ECERS-R and CIS as measures of quality.

To examine measurement invariance for the ECERS-R (Harms et al., 1998) and the CIS (1989) across INC and NSN classrooms, two multiple group confirmatory factor analyses were conducted for each instrument. To examine group differences in latent variables, comparisons of latent variables and of derived composite scores were conducted. All analyses were weighted to account for the complex sampling design of the ECLS-B. Because the sampling unit for the ECLS-B was children rather than classrooms, findings from the present study are not nationally representative of classrooms.

In this chapter, findings from the present study are interpreted, implications of the findings are discussed, and recommendations for future research are presented,

followed by a summary of the study. Findings associated each research question are integrated within each section.

Interpretation of Findings

Findings are considered with respect to previous research conducted using the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989). In addition to considering findings related to measurement invariance of each instrument, evidence for the measurement models proposed in the present study is interpreted with respect to findings from previous studies examining the factor structures of the ECERS-R and the CIS. All findings pertaining to group differences in latent variables were interpreted based on statistical significance at the .05 threshold as well as the magnitude of effect sizes. As recommended by Wilkinson and the Task Force on Statistical Inference (1999) and emphasized by Thompson (2002; 2007), effect sizes from the present study were interpreted with respect to effect sizes from other studies examining differences in ECERS-R scores across INC and NSN classrooms. It was not possible to make direct comparisons, however, because previous researchers typically have reported findings based on published subscale scores rather than empirically derived latent variables or composite scores. Therefore, findings were also interpreted with respect to benchmarks for large (.8), moderate (.5) and large (.2) effects recommended by Cohen (1992). In the next two sections, an interpretation of findings pertaining to the measurement invariance of each instrument is presented, followed by an interpretation of findings related to group differences on latent variables and derived composite scores.

Measurement Invariance of the ECERS-R

Measurement invariance for the ECERS-R was examined using a two-step process. First, a configural invariance model was tested to confirm that the factor structure of the proposed three-factor model fit the data from INC and NSN classrooms adequately and that the factor structure was equivalent across groups. Second, the strict invariance model was tested to examine the extent to which factor loadings, item thresholds, and residual variances were equivalent across groups. The RMSEA, CFI, and TLI indices indicated good model fit for the configural invariance model, suggesting the ECERS-R provides a measure of three latent variables related to the quality of the provision of activities and materials, language and interactions, and personal care and safety routines in INC and NSN classrooms. Model fit indices for the configural invariance model also indicated the items associated with these variables were equivalent across INC and NSN classrooms. The RMSEA, CFI, and TLI indices for the strict invariance model also indicated good model fit. Given the restrictions imposed on the strict invariance model (i.e., equality of factor loadings, item thresholds, and residual variances across groups), these findings provide strong support for measurement invariance of the ECERS-R across INC and NSN classrooms.

Internal consistency score reliabilities for each of the derived subscales were high and similar across INC and NSN classrooms, suggesting similar inter-item correlations across groups. An examination of the proportions of items scores in INC and NSN classrooms revealed similar patterns in score responses across the two groups of classrooms, providing further validation for the equivalence of item thresholds across groups.

Taken together, findings from the configural and strict invariance models tested in the present study provide preliminary evidence to suggest (a) the ECERS-R measures three latent variables related to the quality of the provision of materials and activities, language and interactions, and personal care and safety routines in INC and NSN classrooms; (b) these three latent variables are characterized by the same ECERS-R items in both types of classrooms; and (c) latent variable scores are comparable across INC and NSN classrooms.

Some limitations in the generalization of these findings should be noted, however. Although Taylor Series linearization was used to account for the clustered sampling approach in the ECLS-B, the classrooms sampled for the study are not representative of all INC and NSN classrooms. Because the ECLS-B was a large, national study, it was not possible to balance raters across regions; rather, raters were assigned to conduct CCOs by region. In the present study, no methods were used to account for rater effects that might have led to an overestimation of measurement invariance. In addition, the present study does not provide any information about ECERS-R items related to provisions for parents and staff, because these items were not administered for the ECLS-B. Had these items been included in the analyses, the number of latent variables might have been different. At the very least, the item composition of the three latent variables identified in the present study would have been different, which might have affected measurement invariance.

Despite the previously noted limitations, the present study provides important validity evidence for the ECERS-R (Harms et al., 1998). It is among the first studies in which measurement invariance for a measure of quality in ECE has been examined; no

other studies were identified that examined assumptions related to measurement invariance of the ECERS-R across different types of ECE classrooms or programs. Although items on the ECERS-R were designed to measure dimensions of quality that would theoretically and logically be expected to be present in any ECE classroom or program, the present study is the first known study to provide empirical evidence in support of measurement invariance of the ECERS-R across two types of ECE classrooms.

Factor Structure of the ECERS-R

Findings from the present study replicated previous findings regarding the factor structure of the ECERS-R (Harms et al., 1998). Namely, there was evidence to suggest a more parsimonious internal structure than that originally proposed by the authors of the instrument. A three-factor structure similar to the one proposed by Gordon et al. (2013) was supported using a confirmatory model across INC and NSN classrooms. An examination of factor loadings for each of the three latent variables proposed in the three-factor model revealed strong and statistically significant associations between items and latent variables, indicating the items are good indicators of the latent variables.

The item compositions for the Activities and Materials factor and the Language and Interactions factor in the present study were similar to those of similarly named factors proposed by other researchers who have examined the factor structure of the ECERS-R (see Table 2-4). Generally, items related to activities and materials were indicators of structural quality, as defined by Cassidy, Hestenes, Hansen, et al. (2005) and items related to the Language and Interactions factor were indicators of process quality (see Figure 1-1). In addition, the internal consistency score reliabilities for these

factors were similar to the Activities/Materials factor and the Language/Interaction factor reported by Cassidy, Hestenes, Hegde, et al. (2005). Findings across studies examining the factor structure of the ECERS-R suggest the instrument measures, at a minimum, aspects of structural and process quality related to the provision of developmentally appropriate materials and activities and basic interactions in the classroom.

In contrast to some earlier studies examining the factor structure of the ECERS-R, the model proposed for the present study retained items related to personal care, health, and safety routines. Retaining these items was preferred for the present study, because they provide information about practices that are not captured on other measures of quality in ECE currently available, but that are reflected in licensing and accreditation standards (e.g., NAEYC, 2013). Typically, these items have been not been included in structural interpretations as part of principal components or exploratory analyses because they did not have strong associations with the obtained factors (Clifford et al., 2005; Cassidy, Hestenes, Hegde et al., 2005). In other studies, these items have been associated with factors that appeared to measure primarily the quality of language and interactions (Perlman et al., 2004; Sakai et al., 2003). One possible explanation for these latter findings is that many of the indicators for these items are designed to capture the extent to which personal care routines are used as an opportunity for more interaction between staff and children (e.g., meals and snacks are times for conversation; pleasant staff-child interactions during toileting; staff explain reasons for safety rules; Bruder and Brant, 1995; Harms et al., 1998).

Another possible explanation for differences in factor composition and the number of items retained across studies examining the factor structure of the ECERS-R (Harms et al., 1998) is some items on the ECERS-R might be measuring multiple dimensions of quality. For example, Item 16 (i.e., Encouraging Children to Communicate), includes indicators related to the types of materials available to encourage children to communicate (e.g., materials that encourage children to communicate are accessible in a variety of interest centers) in addition to indicators related to the quality of staff-child interactions (e.g., staff balance listening and talking appropriately for age and abilities of children during communication activities; Harms et al., 1998). In the present study, factor correlations for the ECERS-R were fairly high, indicating items might load onto multiple factors. Factor correlations in the present study were similar to those noted by Gordon et al. (2013) and Perlman et al. (2004). Although these findings are similar across studies, interpretations regarding the extent to which the ECERS-R provides a multidimensional measure of quality differ. In addition to reporting high inter-factor and inter-item correlations, Perlman et al. noted the factor related to the provision of activities and materials accounted for more than 70% of the common variance in item scores and proposed the ECERS-R is a unidimensional measure of quality in ECE. Gordon et al. also reported relatively high factor loadings for all items in a one-factor model, but rejected the one-factor model in favor of a three-factor model based on model fit indices and interpretability of latent variables. Although items on the ECERS-R do capture multiple aspects of quality, an in-depth review of item indicators revealed that, within items, indicators can generally be described either as predominately capturing structural quality or process quality

(Cassidy, Hestenes, Hansen, et al., 2005). Items for which there is no clear pattern of indicators being predominately related to structural or process quality are generally those items related to personal care and safety routines, which might explain mixed findings regarding the extent to which these items represent a distinct latent variable.

Across studies reviewed as part of the literature review for the present study, fewer factors than those reflected in the published subscales for the ECERS-R (Harms et al., 1998) were supported empirically. A number of researchers have advocated the use of shortened scales consisting primarily of items from the originally published Activities, Language, and Interactions subscales (Cassidy, Hestenes, Hegde, et al., 2005; Clifford et al., 2005; Early et al., 2006; Hestenes et al., 2008) and have begun to use latent variable subscale scores as measures of structural and process quality rather than using the published subscale scores. Despite these recommendations, the majority of studies examining differences in dimensions of quality in ECE across different types of ECE classrooms or programs have continued to report only the published total and subscale scores. Findings from the present study provide evidence to support comparisons of three latent variable scores related to the provision of developmentally appropriate activities and materials, language and interactions, and personal care and safety routines across INC and NSN classrooms. Given the findings of measurement invariance for each instrument examined in the present study, group comparisons of latent variable scores and derived composite scores were conducted.

Group Differences in ECERS-R Latent Variables

The examination of group differences in the ECERS-R latent variables identified in the present study was conducted using two different analytical procedures. First, differences in mean factor scores were examined by setting the factor means and

variances for NSN classrooms equal to zero and one, respectively, and allowing the estimation of factor means and variances for INC classrooms in the strict invariance model. Use of these model specifications provided a standardized estimate of mean differences of latent variables. The second procedure used to examine group differences in latent variables involved regression analyses to determine the effect of classroom type on derived composite scores. Findings are discussed relative to findings from previous studies in which comparisons of structural, process, or global quality across classrooms with differing compositions of children with special needs have been made using scores from the ECERS-R.

Two caveats are noted regarding the interpretation of findings from the present study relative to previous studies. First, the way in which inclusive classrooms were defined in the present study differs from the definitions used in previous studies; thus, the samples are not directly comparable to one another. Second, previous studies have not included comparisons based on the same latent variables that were used in the present study. In order to make the most direct comparisons possible, findings regarding subscales and derived composite scores from previous studies that were most comparable in terms of item composition to the latent variables described in the present study were used.

Activities and materials

Findings regarding differences in latent variables and derived composite scores for the Activities and Materials factor revealed statistically significant differences indicating higher quality in INC classrooms compared to NSN classrooms. In the present study, mean differences in latent variables and derived composite scores were small to moderate. Previous findings regarding the magnitude of differences in the

provision of activities and materials across INC and NSN classrooms differ across studies. Hestenes et al. (2008) reported findings from similar analyses with two samples of classrooms. Derived composites for these studies were based on fewer items related to the provision of developmentally appropriate activities and materials ($v = 9$). Effect sizes for the first sample, which included a large number of classrooms with relatively high quality ratings within and across groups, were similar to those found in the present study and were statistically significant. In the second sample, classrooms had more diverse quality ratings but the sample size for the study was substantially smaller. Although no statistically significant differences were indicated, effect sizes for this group of classrooms indicated a mean difference of almost one quarter standard deviation higher than those found in the present study or in the study Hestenes et al. conducted with a much larger sample of classrooms.

Studies in which comparisons were made based on the Activities subscale score proposed by the developers of the ECERS-R also yielded mixed findings regarding the magnitude of effects. Grisham-Brown et al. (2010) found very robust and statistically significant effects indicating higher quality in INC classrooms ($d = 1.15$). Although they were not statistically significant, Knoche et al. (2006) found similar effect sizes to those reported in the present study. The variability in the magnitude of effects across studies might be due to differences in the study samples or to differences in the item sets for which comparisons were made across studies. Despite variability in analytic procedures and the magnitude of effects, there appears to be a trend across studies suggesting the activities and materials provided in classrooms where children with and without special needs are enrolled are of higher quality than those in classrooms where

no children with special needs are enrolled. One possible explanation for this is that teachers in classrooms where children with special needs are enrolled provide a wider variety of materials to meet the developmental needs of all children in the classroom.

Language and interactions

The two statistical approaches used in the present study to examine differences in the Language and Interactions factor yielded different findings. The effect size for differences in latent variables was small to moderate and statistically significant, indicating higher quality in INC classrooms. When derived composite scores were compared across groups, the effect size decreased by approximately one tenth of a standard deviation, and the difference in scores across groups was not statistically significant. One potential reason for this might be that averaging scores across items assumes each item contributes equal weight to the subscale score. Latent variable scores from the strict invariance model took into account the respective association of each item to the latent variable, which might have lead to the slight difference in findings across the two approaches. It is also important to note that, although there were statistically significant differences in latent variable scores, these differences were barely significant using the .05 threshold. Given the large sample size of the present study, it is possible that the statistical significance of the findings is due to increased power resulting from the large sample.

The effects in the present study regarding the Language and Interactions factor were generally more modest than those found by other researchers; however, previous studies have had mixed results regarding the extent to which there are statistically significant differences in practices related to language and interactions as well as the magnitude of differences across INC and NSN classrooms. Hestenes et al. (2008)

found larger effects with respect to their Language/Interactions factor in two samples of classrooms, but these differences were only statistically significant for one sample of classrooms examined.

Grisham-Brown et al. (2010) also reported larger effect sizes similar to those found by Hestenes et al. when they made comparisons of the Interactions subscale score proposed by the developers of the ECERS-R (Harms et al., 1998), but the effect was not statistically significant. Findings from Knoche et al. (2006) paralleled those in the present study most closely, with modest effect sizes that were not statistically significant.

The sample sizes for the majority of the previous studies mentioned were relatively small, which might have resulted in low power to detect differences across INCCs and CNCNs. It is important to note, however, that with the exception of one study (Hestenes et al., 2008), researchers have not reported statistically significant differences in practices related to interactions. Although there appears to be a trend for slightly higher means in INC classrooms, it is possible the ECERS-R does not measure practices related to language and interactions with enough specificity to detect notable differences across INC and NSN classrooms. Evidence of a lack of specificity regarding these practices is apparent in the relatively high scores and limited variability in scores for latent variables and published subscale scores across studies, including the present study.

Personal care and safety

Findings regarding differences in latent variable scores and derived composite scores for the Personal Care and Safety factor also revealed statistically significant differences indicating higher quality in INC classrooms compared to NSN classrooms. In

all four analyses, effect sizes were small to moderate and statistically significant, indicating higher quality in INC classrooms. Effect sizes for both the latent variable scores and the derived composite scores were small to moderate, and were comparable to those found in previous studies.

Although the item sets for the present study differed slightly from those used in previous studies, the effect sizes for the Personal Care and Safety factor were consistent with those for the Personal Care and Routines subscale from studies conducted by Grisham-Brown et al. (2010) and Knoche et al. (2006). The present study is the only study in which statistically significant differences between groups were detected. The most likely explanation for this is that the two previous studies were underpowered to detect statistically significant differences due to substantially smaller sample sizes than the sample size in the present study.

Measurement Invariance of the CIS

Measurement invariance of the CIS (Arnett, 1989) was examined using the same two-step process described for the ECERS-R (Harms et al., 1998). The measurement model tested for the CIS in the present study was a bi-factor model proposed by Colwell et al. (2012). In contrast to the four-factor model proposed by Arnett (1989), the model proposed in the present study consisted of one general factor related to caregiver interactions. Two methodological factors were also included in the model, which were specified to be uncorrelated with the general factor. The methodological factors were included in the model to control for inter-item correlations related to the negative or positive orientation of item wording (Collwell et al., 2012).

The RMSEA, CFI, and TLI indices for the configural invariance model indicated good model fit, suggesting that the CIS provides a global measure of the quality of

caregiver interactions. Model fit indices for the configural invariance model also indicated that the items associated with these variables were equivalent across INC and NSN classrooms. The RMSEA, CFI, and TLI indices for the strict invariance model, which provides strong evidence of measurement invariance, also indicated good model fit.

Internal consistency score reliabilities for the substantive factor (i.e., Caregiver Interactions) and the two methodological factors were high and nearly identical across INC and NSN classrooms. An examination of the proportions of items scores in INC and NSN classrooms revealed similar patterns in score responses across the two groups of classrooms, with both groups of classrooms receiving primarily raw scores of 3 or 4 for every item.

Taken together, these findings provide preliminary evidence to suggest (a) the CIS measures one substantive latent variable related to caregiver interactions and two methodological latent variables related to the positive or negative orientation of item wording in INC and NSN classrooms; (b) these three latent variables are characterized by the same CIS items in both types of classrooms; and (c) latent variable score values are comparable across INC and NSN classrooms, permitting examinations of group differences in latent variable scores. It is important to note, however, that no methods were used in the present study to account for the regional assignment of raters to assess classrooms in the ECLS-B study. As such, it is possible that rater effects might have led to an overestimation of measurement invariance for the CIS. Nonetheless, the present study is one of only a handful of studies in which the internal structure of the CIS has been examined. Given such limited validity evidence based on internal

structure, findings from the present study contribute additional information about the psychometric properties of the CIS.

Factor Structure of the CIS

The present study provided additional evidence to support the bi-factor model for CIS scores proposed by Colwell et al. (2012). Although the analytic samples for the present study and the Colwell et al. study were both drawn from the ECLS-B, findings from the present study reflect application of the bi-factor model to two distinct groups of classrooms, which were not differentiated by Colwell and colleagues. Although previous studies examining the factor structure of the CIS have resulted in a multiple-factor scale, it has been suggested that these findings are likely related to confounding from inter-correlations between items due to the positive or negative orientation of the items (Collwell et al., 2012). For example, all of the items in the Sensitivity subscale originally proposed by Arnett (1989) have a positive orientation, whereas all of the items in the Harshness subscale have a negative orientation. When the orientation of item wording is controlled for using a bi-factor model, there is evidence to suggest the CIS is a unidimensional measure of caregiver interactions.

An examination of factor loadings for the Caregiver Interactions factor revealed strong and statistically significant associations between items and latent variables, indicating that the items are good indicators of the latent variables. Furthermore, internal consistency score reliabilities for this factor were very high, indicating high inter-item correlations. Factor loadings for the methodological factor representing items with negative orientation were generally high as well; however, there were four items that did not appear to be associated strongly with this factor. All of the factor loadings for the Positive Orientation factor were large and statistically significant, indicating strong

associations with the latent variable. Although findings regarding the methodological factors are reported, it is important to note that only the general factor related to caregiver interactions is of substantive interest. The use of this scale score has been used in previous studies to examine differences in caregiver interactions across INC and NSN classrooms (Knoche et al., 2006; Wall et al., 2006). The present study is the first study to provide empirical evidence to suggest measurement invariance of this latent variable. Given such evidence, comparisons of latent variable scores and derived composite scores across INC and NSN classrooms were conducted.

Group Differences in CIS Latent Variables

Findings regarding group differences in CIS scores across INC and NSN classrooms were similar in the present study to those reported by Knoche et al. (2006) and Wall et al., (2006). Across all three studies, effect sizes were marginal and were not statistically significant. In the present study, there was extremely limited variability in item-level scores, with most classrooms receiving raw scores of either 3 or 4 for every item. Means for derived composite score means were similar to those reported in previous studies. Given this limited variability, it is likely that CIS scores would not evidence differences in caregiver interactions across the two types of center-based classrooms examined, even if such differences existed.

Implications from the Present Study

The present study highlights the need for empirical evidence regarding measurement invariance of instruments used to measure structural, process, and global ECE classroom program quality. Although there are a number of instruments available to assess the structural, process, or global quality in ECE classrooms and programs, the present study is among the first in which measurement invariance of such

instruments has been examined. For this study, measurement invariance was examined for two instruments that have been used widely to characterize structural, process, and global quality in ECE for numerous purposes across a variety of classroom and program types. Findings from the present study provide additional validity evidence for the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989). Implications of these findings are discussed with respect to practical application of findings and implications for policy and research.

Practical Implications of Findings

Findings from the present study have important implications for the practical application of the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989) as measures of quality in INC and NSN classrooms. The present study provides preliminary empirical evidence to suggest each instrument provides equivalent measures of quality across INC and NSN classrooms. Evidence of measurement invariance suggests meaningful interpretations can be made of comparisons of scores from these instruments across INC and NSN classrooms. It is important to note, however, that the scores for which this evidence exists are latent variable scores derived from the measurement models tested in this study rather than the subscales originally proposed by the authors of the instruments. Findings from this study, combined with findings from previous studies examining the factor structure of the ECERS-R and the CIS, provide support for interpreting latent variable scores measured by these instruments rather than making inferences about global quality by combining information across different variables (e.g., using original subscale scores).

Findings in the present study suggest ECERS-R scores reflect latent variables related to activities and materials (i.e., structural quality) and language and interactions

(i.e., process quality), which have also been evident in previous studies (e.g., Cassidy, Hestenes, Hegde, et al., 2005; Clifford et al., 2005; Gordon et al., 2013; Perlman et al., 2004). It appears from the present study and from previous studies that ECERS-R scores are sensitive to detect variation in the quality of activities and materials provided to children across different types of classrooms (e.g., Grisham-Brown et al., 2010; Hestenes et al., 2008; Knoche et al., 2006; Wall et al., 2006).

Although there was evidence of measurement of a latent variable related to language and interactions in the classroom, a review of the proportions of item scores for the items characterizing this variable revealed limited variability in item score responses. Within and across INC and NSN classrooms, the highest percentage of classrooms received scores of either 6 or 7. Similar limitations in the variability of quality scores related to language and interactions on the ECERS-R have been reported in previous studies (e.g., Grisham-Brown, 2010; Hestenes et al., 2008; Knoche et al., 2006; Wall et al., 2006).

In addition, findings regarding the CIS (Arnett, 1989) were similar to those of the Language and Interactions factor for the ECERS-R (Harms et al., 1998). Although there was evidence to suggest the CIS provides a global measure of caregiver interactions, variability of scores was extremely limited, with the vast majority of classrooms receiving scores of 3 or 4 for all items. Previous researchers have reported similar restrictions in the range of CIS scores (Colwell et al., 2012; Knoche et al., 2006; Wall et al., 2006).

Taken together, findings regarding the ECERS-R Language and Interaction factor and the CIS Caregiver Interaction factor suggest it might not be appropriate to use these instruments to examine differences in process quality across different types

of ECE classrooms (e.g., INC and NSN). Alternative approaches (e.g., observation coding system) or instruments (e.g., Classroom Assessment Scoring System; Pianta et al., 2008) might be needed to detect meaningful variations in process quality across these types of classrooms, should such differences exist.

In addition to the two latent variables often described in the literature, the present study provided additional evidence that the ECERS-R (Harms et al., 1998) measures a latent variable related to personal care and safety routines. The item composition for the Personal Care and Safety factor identified in the present study and by Gordon et al. (2013) is very similar to the Personal Care Routines subscale published by the developers of the ECERS-R. Items characterizing this factor have been identified as important indicators of ECE global quality, because they relate to basic functions of ECE programs and classrooms that ensure the health and safety of children in attendance (Cryer, 1999). Despite their importance to assessing the overall quality of an ECE classroom or program, these indicators are not easily categorized as being related to either structural or process quality using current conceptualizations of quality in ECE. Previous studies provide evidence of this assertion given mixed findings about whether or which latent variable or quality dimension these ECERS-R items were associated with (Cassidy, Hestenes, Hansen et al., 2005; Cassidy, Hestenes, Hegde, et al., 2005; Clifford et al., 2005; Perlman et al., 2004; Sakai et al., 2003). The ECERS-R is one of few instruments designed to capture indicators of quality related to personal care routines and health and safety practices, as most instruments focus primarily on specific aspects of either structural or process quality (see Table 2-1; Snow & Van Hemel, 2008). Furthermore, there was evidence in the present study to suggest the

ECERS-R is sensitive to detect differences in practices associated with this variable across different types of classrooms and the magnitudes of these effects across INC and NSN classrooms was the most stable and comparable to those found in previous studies. Given the importance of these indicators, it might be prudent to consider alternate conceptualizations of quality in which the promotion of the health and safety of children is distinguished as a separate dimension of quality observed in ECE classrooms or programs rather than trying to characterize them as being indicative of structural or process quality (Burchinal et al., 2011).

In summary, findings from the present study suggest that although there is evidence of measurement invariance for both ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989) across INC and NSN classrooms, neither instrument is particularly sensitive to detect variation in process quality across these two types of classrooms, if such differences exist. Previous researchers have considered the ECERS-R as a measure of process quality; however, findings from the present study suggest it is most useful as a measure of structural quality, particularly when the contemporary definitions of structural and process quality provided by Cassidy, Hestenes, Hansen, et al. (2005) are applied. The ECERS-R also appears to be a useful tool for measuring variations in practices associated with personal care routines and the provisions for the health and safety of children. Of particular note is the importance of interpreting scores based on empirically validated latent variable subscales rather than the published subscales.

Policy and Research Implications of Findings

Findings from present study have implications for policy decisions made based on scores from either the ECERS-R (Harms et al., 1998) or the CIS (Arnett, 1989).

Of particular importance are the implications for the implementation of quality rating and improvement systems (QRIS). QRIS are designed to quantify dimensions of quality in ECE programs for both program accountability and improvement purposes and to communicate quality ratings to consumers (Tout et al., 2009; Tout et al., 2010). According to recent data, 37 states have launched a statewide QRIS, three states have completed a pilot of a QRIS, two states have implemented regional QRIS, and seven states are in planning stages of implementing a QRIS (QRIS National Learning Network/Build Initiative, 2013). Although there is variability in the implementation of QRIS across states, a common feature of QRIS is a process for monitoring quality standards. In a review of 26 QRIS implemented nation-wide, Tout et al. (2010) found the ECERS-R was the predominant observational instrument used to determine environmental quality ratings in center-based care settings for preschool-aged children. The most common method for assigning quality ratings was based on cut-points of the overall ECERS-R score across the classrooms samples for the program. Although similar methods have commonly been used to describe “global” quality in ECE research, this practice does not align with contemporary conceptualizations of focusing on dimensions of ECE quality (i.e., process quality, structural quality), nor is it aligned with accumulated evidence regarding the factor structure of the ECERS-R.

Few states have engaged in systematic processes for evaluating the extent to which decisions about program quality standards and associated measurement strategies result in accurate or meaningful quality ratings (Tout & Starr, 2013; Zellman & Fiene, 2012). Given the potentially high stakes involved for states, for all ECE programs, and for individual ECE programs, validation of the measures used in QRIS

systems and the processes used to assign quality ratings are very important. Recent guidance highlights the importance of examining the psychometric properties of the measures used to assess quality and evaluating the quality of outputs of the rating process, which includes examining the extent to which there is meaningful variation of ratings within and across programs (Tout & Starr, 2013; Zellman & Fiene, 2012). Given evidence from the present study and other recent studies suggesting a lack of sensitivity in ECERS-R scores for detecting variation in process quality, it might be useful to consider alternate measures for this purpose.

Findings from the present study are also relevant for the design and implementation of large-scale evaluations of ECE classroom or program quality, which tend to involve the aggregation of scores across various program and classrooms types. Preliminary evidence from this study suggests measures of structural and process quality yielded by the ECERS-R and CIS are invariant across two common types of ECE classrooms included in samples for large-scale evaluations: INC and NSN classrooms. The measurement models validated in the present study also provide an alternate structure for summarizing scores on the ECERS-R and the CIS. Furthermore, the factor structures identified in the present study are generally aligned with those reported by previous authors (e.g., Cassidy, Hestenes, Hegde, et al., 2005; Clifford et al., 2005; Gordon et al., 2013; Perlman et al., 2004).

An additional contribution of the present study is evidence to support comparisons of derived composites across INC and NSN classrooms. Since the passage of P.L. 99-457, there has been rising interest in characterizing quality in ECE classrooms where children with special needs are enrolled in order to gain a better

understanding of the implementation of inclusive practices (Buysse et al., 1999). Previous researchers have questioned the extent to which it is appropriate to use instruments validated primarily in ECE programs for children without special needs to characterize dimensions of quality in preschool ECE classrooms in which children with special needs are enrolled (Bailey et al., 1982; Buysse et al., 1999; Soukakou, 2012; Spiker, Hebbeler, & Barton, 2011). Despite these cautions, both the ECERS-R and the CIS have been used in a number of studies to make comparisons of structural, process, and global quality across ECE classrooms and programs with differing compositions of children with special needs. The present study is the first study identified that provides evidence to support comparisons of ECERS-R and CIS scores across different types of classrooms. It is important to note, however, the evidence of measurement invariance for the ECERS-R and the CIS pertains only to the models tested in the present study, and cannot be generalized to alternate measurement models.

The present study is also one of only a few studies in which comparisons of quality have been made across INC and NSN classrooms using scores from empirically validated latent variable subscales. Findings in the present study suggest higher quality practices related to personal care routines, the promotion of children's health and safety, and the provision of developmentally appropriate materials and activities. Although these findings provide additional information about the quality in ECE classrooms in which children with special needs are enrolled, they do not provide information about the quality of inclusive practices, nor do they provide information about the quality of care experienced directly by children with special needs. In

addition, the extent to which these findings are meaningful with respect to child outcomes has not yet been determined.

Recommendations for Future Research

A number of recommendations for future research are noted. First, there is a need for more comprehensive validity studies utilizing analytical procedures drawn from both classical test theory and more advanced methodologies for measures of quality in ECE classrooms or programs (Bryant et al., 2011; Gordon et al., 2013). The present study is one of only two studies identified in which measurement invariance has been examined for measures of quality in ECE, and there have been very few studies conducted in which more advanced methodologies, such as IRT models, have been used to establish validity evidence for such measures. Although the present study provides preliminary validity evidence regarding measurement invariance of the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989), findings are limited to INC and NSN classrooms. Additional research is needed to determine the extent to which there is evidence of measurement invariance of these instruments for other types of ECE classrooms for both the instruments of interest in the present study and for other measures designed to measure dimensions of quality in ECE. Such evidence is important for making meaningful inferences about the quality of ECE in large, heterogeneous samples of ECE classrooms, for making comparisons of quality across different types of ECE classrooms, and for examining relationships between the quality in ECE and child outcomes.

The need for empirical studies establishing measurement invariance for instruments administered in ECE classrooms is not limited to measures of quality in ECE classrooms or programs. In the last 30 years, there has been a growing body of

evidence suggesting a positive relationship between the quality in ECE and child outcomes, which has resulted in heightened awareness of the importance of high quality ECE for all children. The impact of high quality ECE for children with environmental or biological vulnerabilities has been especially influential for policymakers, practitioners, and researchers.

Although the impact of the quality of ECE on child outcomes has been accepted widely, a meta-analysis of the associations between the quality of ECE and child outcomes revealed only modest associations between these two variables (Burchinal et al., 2009; Burchinal et al., 2011). One possible explanation for this finding is the need for more refined measures of various dimensions of quality in ECE classrooms and programs; however, there is also a necessity to conduct more rigorous studies to establish validity evidence, including evidence of measurement invariance, for child outcome measures across different groups of children. Although such studies are more prevalent for child outcome measures than the quality of ECE, these studies are typically focused on establishing measurement invariance across genders or across longitudinal administrations of the instrument. Additional studies are needed to examine the extent to which measures of child outcomes are invariant across diverse groups of children, including children with and without special needs.

In addition to conducting more comprehensive validity studies for existing measures of the quality of ECE and child outcome measures, there is a need for the development of measures that are more sensitive to variations in specific dimensions of quality across ECE classrooms and programs and for individual children. As noted previously, one possible explanation for the modest relationships between the quality of

ECE and children's developmental outcomes noted by Burchinal et al. (2009; 2011) is a lack of specificity and sensitivity of existing measures of quality in ECE classrooms and programs, particularly with respect to process quality. Process quality is cited as having the most influential effect on child outcomes, yet evidence from the present study and from previous studies suggests that global rating scales such the ECERS-R and the CIS, which are commonly used to assess aspects of process quality, might not be sensitive enough to detect variability in process quality. Furthermore, the unit of analysis for the majority of existing measures is the classroom rather than individual children. The development and validation of instruments that can be used reliably to detect variations in specific dimensions quality across a variety of ECE classrooms and for individual children will be important to gather more specific descriptive information about dimensions of quality in ECE and the relationship between these dimensions and children's developmental outcomes, particularly for children with special needs.

A final recommendation for future research pertains to studying the quality of ECE in classrooms where children with special needs are enrolled. A number of studies have been conducted to provide descriptive information about the ECE classrooms and programs in which children with special needs are enrolled. Although these studies provided important information about the quality of classrooms in which children with special needs were enrolled compared to classrooms in which no children with special needs were enrolled, the extent to which these findings are meaningful with respect to developmental outcomes for children with special needs is unknown. Despite a growing interest in the impact of quality of ECE on developmental outcomes for children with environmental vulnerabilities (e.g., low income), no known studies have been conducted

to examine such relationships for children with special needs. This line of research is, of course, dependent on the availability of measures of quality in ECE classrooms and programs as well as child outcomes that are technically adequate for detecting such relationships. Nevertheless, it is an important line of research that will help researchers determine the extent to which observed differences in the quality of ECE across classrooms with differing compositions of children with special needs is actually meaningful for the children with special needs who are enrolled in these classrooms.

Summary

The quality of ECE and dimensions of quality has become a topic of major concern for researchers, policymakers, and early childhood practitioners. Findings from several studies in which different dimensions of quality in ECE have been related to children's developmental outcomes have suggested high-quality ECE experiences are related positively to children's short- and long-term developmental outcomes (e.g., Burchinal et al., 2008; NICHD ECCRN 1998, 2002, 2003; Peisner-Feinber & Burchinal, 1997; Peisner-Feinberg et al., 2001). Findings from these studies, along with the adoption of early childhood policies and recommended practices supporting the education of children with special needs in inclusive ECE environments has spurred a growing body of research to characterize the quality of ECE in center-based ECE preschool programs or classrooms where young children with special needs are enrolled. Both the ECERS-R (Harms et al., 1998) and the CIS have been used for this purpose in a variety of classroom types (including classrooms in which children with special needs are enrolled) and to make comparisons of quality across ECE classrooms and programs with differing compositions of children with special needs. Although these instruments were designed to capture elements of quality that should be present

in any type of ECE classroom, these assumptions had not been tested empirically prior to the present study.

The primary goal of the present study was to determine the extent to which measures of quality commonly used in ECE, the ECERS-R (Harms et al., 1998) and the CIS (Arnett, 1989), were invariant across a large sample of INC and NSN classrooms observed as part of the ECLS-B. A secondary goal of the present study was, in the event there was evidence to suggest measurement invariance of the instruments, to determine if there were differences in latent variables and derived composite scores for each instrument across INC and NSN classrooms. Secondary analyses were conducted with cross-sectional data from 4-year data collection wave of the ECLS-B. Multiple group confirmatory factor analyses were conducted separately for each instrument to examine measurement invariance and to examine group differences in latent variables. Regression analyses were conducted to examine group difference in derived composite scores.

Findings from the multiple group confirmatory factor analyses provided strong evidence of measurement invariance for both the ECERS-R (Harms et al., 1998) and the CIS (1989). The measurement models tested in the present study were derived from previous research findings focused on the factor structure of the ECERS-R and the CIS. For both instruments, more parsimonious measurement models than those proposed by the instrument developers were tested and validated. Although there was evidence to suggest each instrument measured a latent variable related to process quality, neither measure appeared sensitive to detect variation in practices associated with process quality within or across INC and NSN classrooms, if such differences exist

Group comparisons of latent variables and derived composite scores suggested higher quality practices associated with the provisions of developmentally appropriate activities and materials, personal care routines, and health and safety practices in INC classrooms compared to NSN classrooms. In general, the magnitude of effects were similar to those found by previous researchers and were in the small to moderate range. Although these effects were statistically significant, the extent to which the magnitude of the effects is meaningful with respect to children's developmental outcomes was not explored in the present study.

The present study highlights the need for more comprehensive validity evidence for instruments used to measure different dimensions of quality in a variety of ECE classrooms and programs. Findings from this study provided additional validity evidence for two such instruments, which have been used in ECE classrooms with differing compositions of children with special needs. The practical, policy, and research implications of these findings were discussed, followed by recommendations for future research regarding the measurement of quality in ECE and its impact on children's developmental outcomes.

APPENDIX A
INSTRUMENT ITEMS AND MEASUREMENT MODELS

Table A-1. Subscales, Items, and Proposed Factors for the ECERS-R

Subscale	Item	Proposed Factor
Space and Furnishings	Indoor Space	Activities/Materials
	Furnishings for routine care and learning	Activities/Materials
	Furnishings for relaxation	Activities/Materials
	Room arrangement	Activities/Materials
	Space for privacy	Activities/Materials
	Child display	Activities/Materials
	Space for gross motor activities	Personal Care/Safety
	Gross motor equipment	Activities/Materials
	Greeting and departure	Language/Interactions
	Meals and snacks	Personal Care/Safety
Personal Care Routines	Nap/rest	Personal Care/Safety
	Toileting/diapering	Personal Care/Safety
	Health practices	Personal Care/Safety
	Safety practices	Personal Care/Safety
	Books and pictures	Activities/Materials
Language and Reasoning	Encouraging communication	Language/Interactions
	Using language to develop reasoning skills	Language/Interactions
	Informal use of language	Language/Interactions
	Fine motor	Activities/Materials
	Art	Activities/Materials
Activities	Music and movement	Activities/Materials
	Blocks	Activities/Materials
	Sand and water	Activities/Materials
	Dramatic play	Activities/Materials
	Nature/Science	Activities/Materials
	Math	Activities/Materials
	Use of TV, video, computers	Activities/Materials
	Promoting acceptance of diversity	Activities/Materials
	Supervision of gross motor activities	Language/Interactions
Interaction	General supervision	Language/Interactions

Table A-1 Continued.

Subscale	Item	Proposed Factor
Interaction	Discipline	Language/Interactions
	Staff-child interactions	Language/Interactions
	Interactions among children	Language/Interactions
Program Structure	Schedule	Activities/Materials
	Free play	Activities/Materials
	Group time	Activities/Materials
	Provisions for children with disabilities	Language/Interactions

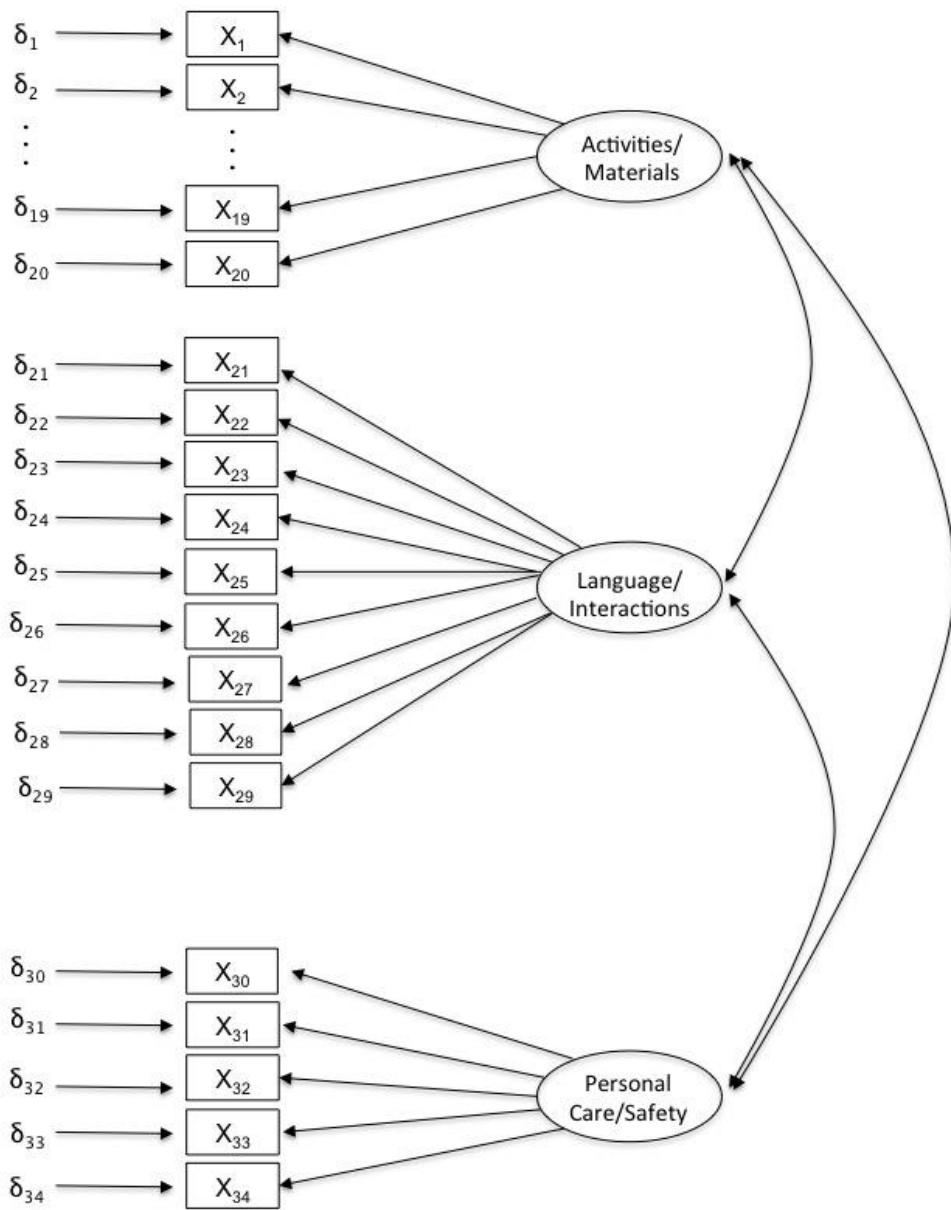


Figure A-1. Measurement Model for the ECERS-R.

Table A-2. Arnett Scale of Caregiver Behavior Items and Item Wording in the ECLS-B Child Care Observations

Subscale	Item	Original Wording	ECLS-B Item wording
Sensitivity	1	Speaks warmly to the children	Speaks warmly to the children (e.g., positive tone of voice, body language)
	3	Listens attentively when children speak to her	Listens attentively when children speak to her (e.g., looks at children, nods, rephrases their comments, engages in conversations)
	6	Seems to enjoy the children	Seems to enjoy the children (e.g., conveys warmth by smiling, touching, taking children's conversations seriously)
	7	When children misbehave, explains the reason for the rule they are breaking	When children misbehave, explains the reason for the rule they are breaking (e.g., discusses consequences, redirects behavior, discusses what to do instead)
	8	Encourages children to try new experiences	Encourages children to try new experiences (e.g., suggests children do it together, helps children start, introduces new materials)
	11	Seems enthusiastic about the children's activities and efforts	Seems enthusiastic about the children's activities and efforts (e.g., congratulates children, states appreciation of their efforts)
	14	Pays positive attention to the children as individuals	Pays positive attention to the children as individuals (e.g., speaks to individual children, uses their names, calls attention to pro-social behaviors, comments on their strengths)
	16	Talks to the children on a level they can understand	Talks to the children on a level they can understand (e.g., uses terms familiar to children, checks for clarification)
	19	Encourages children to exhibit pro-social behavior, e.g., sharing, cooperating	Encourages children to exhibit pro-social behavior (e.g., sharing, cooperating, pairs socially skillful with those children that need practice)
	25	When talking to children, kneels, bends, or sits at their level to establish better eye contact	When talking to children, kneels, bends, or sits at their level to establish better eye contact (e.g., ensures a connections when having a conversation)

Table A-2 continued

Subscale	Item	Original Wording	ECLS-B Item wording
Harshness	2 ^a	Seems critical of the children	Seems critical of the children (e.g., puts children down, uses sarcasm)
	4 ^a	Places high value on obedience	Places high value on obedience (e.g., expects children to follow and adult agenda, fails to respond to daily events in a flexible way)
	10 ^a	Speaks with irritation or hostility to the children	Speaks with irritation or hostility to the children (e.g., sharp tone, raises voice)
	12 ^a	Threatens children in trying to control them	Threatens children in trying to control them (e.g., uses bribes and threats of punishment)
	17 ^a	Punishes the children without explanation	Punishes the children without explanation (e.g., does not discuss infraction)
	20 ^a	Finds fault easily with children	Finds fault easily with children (e.g., negative tone, critical)
	22 ^a	Seems to prohibit many of the things children want to do	Seems to prohibit many of the things children want to do (e.g., adheres to rigid schedule or adult outcomes agendas)
	26 ^a	Seems unnecessarily harsh when scolding or prohibiting children	Seems unnecessarily harsh when scolding or prohibiting children (e.g., angry tone, shakes children, uses physical punishment, uses "time out" without explanation)
	24	Expects the children to exercise self-control, e.g., to be undisruptive for group, teacher-led activities, to be able to stand in line calmly	Expects the children to exercise a reasonable amount of self-control (e.g., expects children to be undisruptive for short group, teacher-lead activities; to be able to stand in lane calmly; reminds children of expectations; and asks for cooperation in supportive ways)
	5 ^a	Seems distant or detached from children	Seems distant or detached from children (e.g., sits apart, does not touch children, does not greet children)
Detachment	13 ^a	Spends considerable time in activity not involving interaction with the children	Spends considerable time in activity not involving interaction with the children (e.g., does adult tasks during child activity periods)

Table A-2 continued.

Subscale	Item	Original Wording	ECLS-B Item wording
Detachment	21 ^a	Doesn't seem interested in children's activities	Fails to show interest in the children's activities (e.g., removes self from children's activities, doesn't talk to children or extend their conversations)
	23 ^a	Doesn't supervise the children very closely	Fails to supervise the children very closely (e.g., withdraws during activities, fails to foresee and forestall mishaps)
Permissiveness	9 ^a	Doesn't try to exercise much control over the children	Exercises too much control over the children (e.g., doesn't take child input, rigid adherence to rules and schedules)
	15 ^a	Doesn't reprimand children when they misbehave	Reprimands children too strongly when they misbehave (e.g., is punitive, fails to acknowledge difficulties of learning self-control, fails to redirect behavior)
	18	Exercises firmness when necessary	Exercises firmness when necessary (e.g., clear directions, checks for understanding)

Note. Adapted from "Early Childhood Longitudinal Study, Birth Cohort (ECLS-B), Preschool Year Data File User's Manual (2005-2006)" by K. Snow et al., 2007, pp. 98-100.

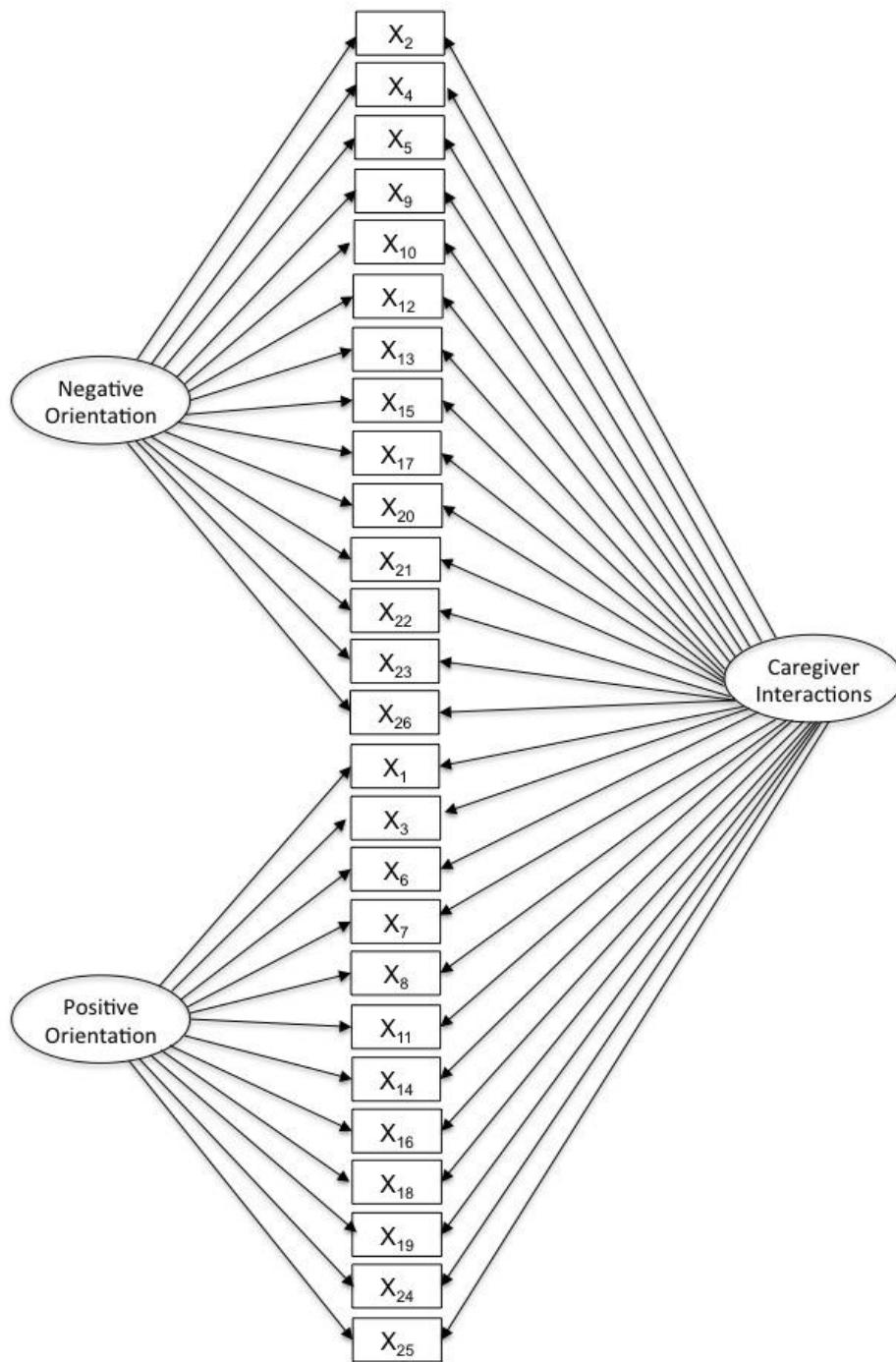


Figure A-2. Bi-factor Measurement Model for the CIS.

APPENDIX B
INTERVIEW QUESTIONS USED TO IDENTIFY CHILDREN WITH SPECIAL NEEDS

Table B-1. Variables from 9-Month Parent Interview Used to Identify Children with Special Needs

Variable	Related Parent Interview Question	Coding
Diagnosed Condition at 9 months of age	CH165 Has a doctor ever told you that {CHILD/TWIN} has the following conditions? (a) Blindness (P1BLIND) (b) Difficulty seeing, including nearsightedness and farsightedness (P1SEE) (c) Difficulty hearing or deafness (P1HEAR) (d) A cleft lip or palate (P1CLEFT) (e) A heart defect (P1HEART) (f) Failure to thrive (P1THRIVE) (g) A problem with mobility or using {his/her} legs to get around (P1MOBIL) (h) A problem with using {his/her} arms or hands (P1HANDS) (i) Down Syndrome (X1SYNDRM) (j) Turner's Syndrome (X1SYNDRM) (k) Spina Bifida (X1SYNDRM) (l) Any other types of special needs or limitations (P1OTSPND)	1: Yes ^a 0: No
Early intervention services at 9 months of age	(X1PRGCMB) ^b CH180 Is {CHILD/TWIN} currently participating in an early intervention program or regularly receiving any services for (his/her) condition(s) from... (a) Your local school district? (b) A state or local health agency? (c) A social service agency? (d) A private doctor's office? (e) A clinic? (f) Some other Source?	1: Yes 0: No

^aIf the parent responded yes for any condition listed in the parent interview question, the composite variable was coded as "yes."

^bComposite variable was provided in the ECLS-B data file and was based on responses to items CH180 a-f.

Table B-2. Variables from 2-Year Parent Interview Used to Identify Children with Special Needs

Variable	Related Parent Interview Question	Coding
Diagnosed condition at 2 years of age	CH180 ^a Has a doctor ever told you that {CHILD/TWIN} has the following conditions? Does {he/she} have.... (a) Blindness? (P2BLIND) (b) Difficulty seeing, including nearsightedness or farsightedness, correctable?(P2SEE) (c) Difficulty hearing or deafness? (P2HEAR) (d) A problem with mobility such as cerebral palsy? (P2MOBIL) (e) A delay in learning to walk? (P2DLYWLK) (f) A delay in learning to talk? (P2DLYTLK) (g) Another developmental delay? (P2DLYOTH) (h) Epilepsy or seizures? (P2EPLPSY) (i) A heart defect? (P2HEART) (j) Mental retardation? (P2MENTL)	1: Yes ^b 0: No
Early intervention at 2 years of age	CH195 Is {CHILD/TWIN} currently participating in an early intervention program or regularly receiving any services for (his/her) condition(s) from... (a) Your local school district? (P2PRGSD) (b) A state or local health or social service agency? (P2PRGHA) (c) A doctor, clinic, or other health care provider? (P2PRGDR) (d) Some other Source? (P2PRGOT)	1: Yes ^b 0: No

^aNot all of the conditions asked about in the parent interview were used to generate composite variables for the present study; only the variables used for the present study are listed.

^bIf the parent responded yes for any condition listed in the parent interview questions, the composite variable was coded as “yes.”

Table B-3. Variables from 4-Year Parent Interview Used to Identify Children with Special Needs

Variable	Related Parent Interview Question	Coding
Evaluated and diagnosed by a professional by 4 years of age	CH181-CH194 Now I have some questions about different disabilities your child might have.	1: Yes ^a 0: No
	Since {CHILD/TWIN} turned 2 years old, has {CHILD/TWIN} been evaluated by a professional in response to {his/her} ability to pay attention or learn? Did you obtain a diagnosis of a problem from a professional? (P3DIAGAT)	
	Since {CHILD/TWIN} turned 2 years old, has {CHILD/TWIN} been evaluated by a professional in response to {his/her} overall activity level? Did you obtain a diagnosis of a problem from a professional? (P3DIAGAC)	
	Since {CHILD/TWIN} turned 2 years old, has {CHILD/TWIN} been evaluated by a professional in response to the use of {his/her} limbs? Did you obtain a diagnosis of a problem from a professional? (P3DIAGLM)	
	Since {CHILD/TWIN} turned 2 years old, has {CHILD/TWIN} been evaluated by a professional in response to {his/her} ability to communicate? Did you obtain a diagnosis of a problem from a professional? (P3DIAGCO)	
	Does {CHILD/TWIN} have you had {CHILD/TWIN}'s hearing evaluated by a professional? Did you obtain a diagnosis of a problem from a professional? (P3DIAGHR)	
	Does {CHILD/TWIN} have you had {CHILD/TWIN}'s vision evaluated by a professional? Did you obtain a diagnosis of a problem from a professional? (P3DIAGVI)	

Table B-3 continued.

Variable	Related Parent Interview Question	Coding
Special education services at 4 years of age	CH195 When a child with a disability or developmental delay receives special education and/or related services sponsored through your local education agency—that is, the school system—these services are initiated after a diagnosis of condition, or evaluation of the child, and the development of an IEP or an IFSP, which is discussed and signed by the parent. Is {CHILD/TWIN} receiving special education services related to an IEP or IFSP?	1: Yes 0: No
Diagnosed condition at 4 years of age	CH200 ^b Has a doctor ever told you that {CHILD/TWIN} has the following conditions? Does {he/she} have.... (a) A problem with mobility such as cerebral palsy? (P3MOBIL) (b) Another developmental delay? (P3DLYOTH) (c) Epilepsy or seizures? (P3EPLPSY) (d) A heart defect? (P3HEART) (e) Mental retardation? (P3MENTL) (f) Autism or PDD? (P3AUTISM) (g) Oppositional Defiant Disorder? (P3ODD) (h) ADHD? (P3ADHD)	1: Yes ^a 0: No

^aIf the parent responded that the child had been diagnosed with any of the conditions listed, the composite variable was coded as "yes."

^bNot all of the conditions asked about in the parent interview were used to generate composite variables for the present study; only the variables used for the present study are listed.

APPENDIX C VARIABLE CODING SYNTAX

SPSS Syntax to Re-code Demographic Data Variables, Program Type, and CCO Variables

*Recode demographic variables

```
DATASET ACTIVATE DataSet1.  
RECODE X3CCOAGE Y3POVRTY Y3SESQ5 (-9=SYSMIS) (-7=SYSMIS) (-  
8=SYSMIS) (-1=SYSMIS).  
EXECUTE.
```

```
RECODE Y1CHRACE (-9=SYSMIS).  
EXECUTE.
```

* Recode ECERS-R and CIS composite variables

```
DATASET ACTIVATE DataSet1.  
RECODE X3ARNDT X3ARNHAR X3ARNPER X3ARNSEN X3ARNTOT X3ECRFD  
X3ECRPC X3ECRLT X3ECRLA X3ECRINT X3ECRPS X3ECRTOT (-9=SYSMIS) (-  
8=SYSMIS) (-7=SYSMIS) (-1=SYSMIS).  
EXECUTE.
```

*Recode ECERS-R item-level variables (item-level variables renamed e1-e37)

```
RECODE e1 e2 e3 e4 e5 e6 e7 e8 e9 e10 e11 e12 e13  
e14 e15 e16 e17 e18 e19 e20 e21 e22 e23 e24 e25 e26 e27 e28 e29 e30 e31 e32  
e33 e34 e35 e36 e37  
(-1=-9) (-8=-9) (-7=-9).  
EXECUTE.
```

*Reverse code negatively oriented CIS items

```
RECODE L3SMSCRT L3OBEDIE L3DISTNT L3EXCONT L3IRRITA L3THREAT  
L3NOTINV L3REPRIM L3PUNWOE L3FINFAU  
L3FALINT L3PROHIB L3FALSUP L3HARSH (1=4) (2=3) (3=2) (4=1).  
EXECUTE.
```

*Recode missing data for CIS items (item-level variables renamed a1-a26)

```
RECODE a1 a2 a3 a4 a5 a6 a7 a8 a9 a10 a11 a12 a13 a14 a15 a16 a17 a18 a19 a20  
a21 a22 a23 a24 a25  
a26 (-1=-9).  
EXECUTE.
```

*Recode ECE provider interview questions related to ECE program type

RECODE type (-1=SYSMIS) (-7=SYSMIS) (-8=SYSMIS) (-9=SYSMIS).
EXECUTE.

SPSS Syntax to Identify Children with Special Needs

*Identify children special needs by 9 months of age via birth weight or condition diagnosed by doctor

Compute lbw=0.
if X1BTHWGT=3 lbw=1.

Compute by_cond_9=0.
if X1SYNDRM=1 OR P1BLIND=1 OR P1SEE=1 OR P1HEAR=1 OR P1CLEFT=1 OR
P1HEART=1 OR P1THRIVE=1
OR P1MOBIL=1 OR P1HANDS=1 OR P1OTSPND=1 by_cond_9=1.

*Identify children identified with special needs by age 2 via EI services or diagnosed condition by doctor

compute svc_2=0.
if X1PRGcmb =1 or P2PRGOT =1 or P2PRGDR = 1 or P2PRGSD =1 or P2PRGHA =1
dis_2=1.

Compute by_cond_2=0.
if by_cond_9=1 or P2BLIND=1 OR P2SEE=1 OR P2HEAR=1 OR P2MOBIL=1 OR
P2DLYWLK=1 OR P2DLYTLK=1 OR P2DLYOTH=1
OR P2EPLPSY=1 OR P2HEART=1 OR P2MENTL=1 by_cond_2=1.

Compute by_bw_cond_svc_2=0.
if lbw=1 or by_cond_2=1 or svc_2=1 by_bw_cond_svc_2=1.

*Identify children identified with special needs by age 4 via EI/Special education services, diagnosed condition by doctor, or evaluation conducted b professional

Compute by_cond_pre=0.
if by_cond_2=1 or P3MOBIL=1 OR P3DLYOTH=1 OR P3EPLPSY=1 OR P3HEART=1
OR P3MENTL=1 OR P3AUTISM=1 OR P3ODD=1 OR P3ADHD=1
by_cond_pre=1.

Compute by_eval_pre = 0.
if P3DIAGAT=1 OR P3DIAGAC=1 OR P3DIAGLM=1 OR P3DIAGCO=1 OR
P3DIAGHR=1 OR P3DIAGVI by_eval_pre=1.

compute svc_pre =0.
if dis_2=1 or P3SPEDU=1 dis_pre=1.

```
compute by_svc_eval_cond_pre=0.  
if svc_pre=1 or by_eval_pre=1 or by_cond_pre=1 by_svc_eval_cond_pre=1.  
compute svcevalcond_pre=0.  
if svc_pre=1 and by_eval_pre=1 and by_cond_pre=1 by_svc_eval_cond_pre=1.  
  
compute all_pre=0.  
if by_svc_eval_cond_pre=1 or lbw=1 all_pre=1.
```

SPSS Syntax to Compute ECE Classroom Type Variables

*Recode ECE provider interview questions related to group size and number of children with special needs

```
RECODE size (-7=SYSMIS) (-8=SYSMIS) (-9=SYSMIS) (-1=0).  
EXECUTE.
```

```
RECODE J3SPECND (-7=SYSMIS) (-8=SYSMIS) (-9=SYSMIS) (-1=0).  
EXECUTE.
```

* Determine number of children without special needs.

```
COMPUTE numtd=size-J3SPECND.  
EXECUTE.
```

* Identify classroom types

```
compute sninc=0.  
if all_pre=1 and numtd >0 sninc=1.
```

```
compute tdnon=0.  
if all_pre=0 and J3SPECND=0 tdnon=1.
```

```
compute tdinc=0.  
if all_pre=0 and J3SPECND >0 tdinc=1.
```

**Syntax to identify proportions children with special needs in classrooms where the ECLS-B focal child had special needs

```
select if (all_pre=1).  
execute.
```

```
DATASET ACTIVATE DataSet1.
```

COMPUTE tot_sn=J3SPECND+1.
EXECUTE.

COMPUTE tot_size=size+1.
EXECUTE.

COMPUTE prpsn = tot_sn / tot_size.
EXECUTE.

compute cltype = -9.
if prpsn < .75 cltype = 1.
if prpsn=.75 cltype=1.

select if cltype=1.
execute.

**Syntax to identify proportion of children with special needs within the sample of classrooms where the ECLS-B focal child did not have special needs, but the ECE provider reported other children with special needs were enrolled

select if (all_pre=0).
execute.

select if (tdinc=1).
execute.

COMPUTE tot_sn=J3SPECND.
EXECUTE.

COMPUTE tot_size=size+1.
EXECUTE.

COMPUTE prpsn=tot_sn / tot_size.
EXECUTE.

compute cltype = -9.
if prpsn <.75 cltype =1.
if prpsn=.75 cltype=1.
execute.

select if cltype=1.
execute.

**Syntax to identify proportion of children with special needs within the sample of classrooms in which neither the ELCS-B focal child nor any other children enrolled in the classroom were identified as having special needs

select if (all_pre=0).
execute.

select if (tdnon=1).
execute.

COMPUTE tot_sn=J3SPECND.
EXECUTE.

COMPUTE tot_size=size+1.
EXECUTE.

COMPUTE prpsn=tot_sn / tot_size.
EXECUTE.

compute cltype = -9.
if prpsn =0 cltype=0.
execute.

select if cltype=0.
execute.

APPENDIX D ANALYTICAL SYNTAX

MPlus Syntax to Test the ECERS-R Configural Invariance Model

```

title: configural invariance;;
data: file is ecersitemswgts-6.5.13.csv;
variable: names are class e1-e37 samwgt psu str;
           usevariables are class e1-e10 e12-e26 e28-e36;
           categorical are e1-e10 e12-e26 e28-e36;
           grouping is class (0=nsn 1=inc);
           stratification is str;
           cluster is psu;
           weight is samwgt;
           missing are all (-9);
analysis: type=complex;
           parameterization=THETA;
           convergence=.00000001;
           iterations=5000;
!following estimates factor loadings and thresholds for cwd and residual
model:   f1 by e1* e2-e6 e8 e15 e19-e26 e28 e34-e36;
           f2 by e9* e16-e18 e29-e33;
           f3 by e7* e10 e12-e14;
           f1 - f3@1;
           model cwd:
           f1 by e1 e2-e6 e8 e15 e19-e26 e28 e34-e36;
           f2 by e9 e16-e18 e29-e33;
           f3 by e7 e10 e12-e14;
           [f1-f3@0];
           f1-f3@1;
!following allows all thresholds to vary across groups;
[e1$1 e2$1 e3$1 e4$1 e5$1 e6$1 e7$1 e8$1 e9$1 e10$1];
[e12$1 e13$1 e14$1 e15$1 e16$1 e17$1 e18$1 e19$1];
[e20$1 e21$1 e22$1 e23$1 e24$1 e25$1 e26$1 e28$1];
[e29$1 e30$1 e31$1 e32$1 e33$1 e34$1 e35$1 e36$1];

[e1$2 e2$2 e3$2 e4$2 e5$2 e6$2 e7$2 e8$2 e9$2 e10$2];
[e12$2 e13$2 e14$2 e15$2 e16$2 e17$2 e18$2 e19$2];
[e20$2 e21$2 e22$2 e23$2 e24$2 e25$2 e26$2 e28$2];
[e29$2 e30$2 e31$2 e32$2 e33$2 e34$2 e35$2 e36$2];

[e1$3 e2$3 e3$3 e4$3 e5$3 e6$3 e7$3 e8$3 e9$3 e10$3];
[e12$3 e13$3 e14$3 e15$3 e16$3 e17$3 e18$3 e19$3];
[e20$3 e21$3 e22$3 e23$3 e24$3 e25$3 e26$3 e28$3];
[e29$3 e30$3 e31$3 e32$3 e33$3 e34$3 e35$3 e36$3];

[e1$4 e2$4 e3$4 e4$4 e5$4 e6$4 e7$4 e8$4 e9$4 e10$4];

```

```

[e12$4 e13$4 e14$4 e15$4 e16$4 e17$4 e18$4 e19$4];
[e20$4 e21$4 e22$4 e23$4 e24$4 e25$4 e26$4 e28$4];
[e29$4 e30$4 e31$4 e32$4 e33$4 e34$4 e35$4 e36$4];

[e1$5 e2$5 e3$5 e4$5 e5$5 e6$5 e7$5 e8$5 e9$5 e10$5];
[e12$5 e13$5 e14$5 e15$5 e16$5 e17$5 e18$5 e19$5];
[e20$5 e21$5 e22$5 e23$5 e24$5 e25$5 e26$5 e28$5];
[e29$5 e30$5 e31$5 e32$5 e33$5 e34$5 e35$5 e36$5];

[e1$6 e2$6 e3$6 e4$6 e5$6 e6$6 e7$6 e8$6 e9$6 e10$6];
[e12$6 e13$6 e14$6 e15$6 e16$6 e17$6 e18$6 e19$6];
[e20$6 e21$6 e22$6 e23$6 e24$6 e25$6 e26$6 e28$6];
[e29$6 e30$6 e31$6 e32$6 e33$6 e34$6 e35$6 e36$6];

e1@1 e2@1 e3@1 e4@1 e5@1 e6@1 e7@1 e8@1 e9@1 e10@1;
e12@1 e13@1 e14@1 e15@1 e16@1 e17@1 e18@1 e19@1;
e20@1 e21@1 e22@1 e23@1 e24@1 e25@1 e26@1 e28@1;
e29@1 e30@1 e31@1 e32@1 e33@1 e34@1 e35@1 e36@1;

output:
tech5;
savedata: DIFFTEST=weighted.original.ci.dat;

```

MPlus Syntax to Test the ECERS-R Strict Invariance Model

```

title: strict invariance;;
data: file is ecersitemswgts-6.5.13.csv;
variable: names are class e1-e37 samwgt psu str;
usevariables are class e1-e10 e12-e26 e28-e36;
categorical are e1-e10 e12-e26 e28-e36;
grouping is class (0=nsn 1=inc);
stratification is str;
cluster is psu;
weight is samwgt;
missing are all (-9);
analysis: type=complex;
parameterization=THETA;
convergence=.00000001;
iterations=5000;
DIFFTEST=weighted.original.ci.dat;

```

```

!following estimates factor loadings and thresholds for cwd and residual
model: f1 by e1* e2-e6 e8 e15 e19-e26 e28 e34-e36;
f2 by e9* e16-e18 e29-e33;
f3 by e7* e10 e12-e14;
f1 - f3@1;
model cwd:

```

f1-f3;

e1@1 e2@1 e3@1 e4@1 e5@1 e6@1 e7@1 e8@1 e9@1 e10@1;
e12@1 e13@1 e14@1 e15@1 e16@1 e17@1 e18@1 e19@1;
e20@1 e21@1 e22@1 e23@1 e24@1 e25@1 e26@1 e28@1;
e29@1 e30@1 e31@1 e32@1 e33@1 e34@1 e35@1 e36@1;

output:

tech5;

savedata: DIFFTEST=weighted.original.strict.dat;

MPlus Syntax to Test the CIS Configural Invariance Model

title: weighted arnett confirmatory configural invariance final;

data: file is recoded_arnettitemswtsonly-6.14.13.csv;

variable: names are class a1-a26 samwgt psu str;

usevariables are class a1-a26 samwgt psu str;

categorical are a1-a26;

grouping is class (0=nsn 1=inc);

stratification is str;

cluster is psu;

weight is samwgt;

missing are all (-9);

analysis: type = complex;

parameterization=THETA;

convergence=.00000001;

iterations=5000;

!following estimates factor loadings and thresholds for cwod and residual

model: interaction by a1* a2-a26;

m1 by a1* a3 a6-a8 a11 a14 a16 a18 a19 a24 a25;

m2 by a2* a4 a5 a9 a10 a12 a13 a15 a17 a20-a23 a26;

interaction @1;

m1-m2@1;

m1 with interaction@0 ;

m2 with interaction@0;

model cwd:

interaction by a1 a2-a26;

m1 by a1 a3 a6-a8 a11 a14 a16 a18 a19 a24 a25;

m2 by a2 a4 a5 a9 a10 a12 a13 a15 a17 a20-a23 a26;

[interaction @0];

[m1-m2 @0];

interaction @1;

m1-m2 @1;

m1 with interaction@0;

m2 with interaction@0;

```

!following allows all thresholds to vary across groups;
[a1$1 a2$1 a3$1 a4$1 a5$1 a6$1 a7$1 a8$1 a9$1 a10$1];
[a11$1 a12$1 a13$1 a14$1 a15$1 a16$1 a17$1 a18$1 a19$1];
[a20$1 a21$1 a22$1 a23$1 a24$1 a25$1 a26$1;

[a1$2 a2$2 a3$2 a4$2 a5$2 a6$2 a7$2 a8$2 a9$2 a10$2];
[a11$2 a12$2 a13$2 a14$2 a15$2 a16$2 a17$2 a18$2 a19$2];
[a20$2 a21$2 a22$2 a23$2 a24$2 a25$2 a26$2 ;

[a1$3 a2$3 a3$3 a4$3 a5$3 a6$3 a7$3 a8$3 a9$3 a10$3];
[a11$3 a12$3 a13$3 a14$3 a15$3 a16$3 a17$3 a18$3 a19$3];
[a20$3 a21$3 a22$3 a23$3 a24$3 a25$3 a26$3;

a1@1 a2@1 a3@1 a4@1 a5@1 a6@1 a7@1 a8@1 a9@1 a10@1;
a11@1 a12@1 a13@1 a14@1 a15@1 a16@1 a17@1 a18@1 a19@1;
a20@1 a21@1 a22@1 a23@1 a24@1 a25@1 a26@1;

```

output:
tech5;
savedata: DIFFTEST=original.weighted.arnett.ci.dat;

MPlus Syntax to Test the CIS Strict Invariance Model

```

title: weighted arnett confirmatory strict invariance final;
data: file is recoded_arnettitemswtsonly-6.14.13.csv;
variable: names are class a1-a26 samwgt psu str;
usevariables are class a1-a26 samwgt psu str;
categorical are a1-a26;
grouping is class (0=nsn 1=inc);
stratification is str;
cluster is psu;
weight is samwgt;
missing are all (-9);
analysis: type = complex;
parameterization=THETA;
convergence=.00000001;
iterations=5000;
DIFFTEST=original.weighted.arnett.ci.dat;
!following estimates factor loadings and thresholds for cwod and residual
model: interaction by a1* a2-a26;
m1 by a1* a3 a6-a8 a11 a14 a16 a18 a19 a24 a25;
m2 by a2* a4 a5 a9 a10 a12 a13 a15 a17 a20-a23 a26;
interaction @1;
m1-m2@1;
m1 with interaction@0;
m2 with interaction@0;

```

```

model cwd:
interaction m1 m2;
[interaction m1 m2];

a1@1 a2@1 a3@1 a4@1 a5@1 a6@1 a7@1 a8@1 a9@1 a10@1;
a11@1 a12@1 a13@1 a14@1 a15@1 a16@1 a17@1 a18@1 a19@1;
a20@1 a21@1 a22@1 a23@1 a24@1 a25@1 a26@1;

output:
tech5;
savedata: DIFFTEST=original.weighted.arnett.strict.dat;

```

SAS™ Syntax to Conduct Group Comparisons of ECERS-R Derived Composites

```

PROC IMPORT OUT= WORK.Crystal
      DATAFILE= "C:\Users\Restrict\Desktop\Crystal\ECLS-B\SPSS Dat
a Files\ecersitemswgts_composite-6.26.13.sav"
      DBMS=SPSS REPLACE;

RUN;
data composite;
set Crystal;
RUN;
QUIT;

proc sort;
by cltype;
RUN;
QUIT;

proc surveyreg;
weight W33P0;
strata W33PSTR;
cluster W33PPSU;
class cltype;
model activities=cltype;
estimate 'mean inclusive' intercept 1 cltype 1;
RUN;
QUIT;

proc surveymeans mean nobs std stderr sum sumwgt min max var;
VAR activities;
by cltype;
weight W33P0;
strata W33PSTR;
cluster W33PPSU;
RUN;

proc surveyreg;
weight W33P0;
strata W33PSTR;
cluster W33PPSU;
class cltype;
model interactions=cltype;

```

```

estimate 'mean inclusive' intercept 1 cltype 1;
RUN;
QUIT;

proc surveymeans mean nobs std stderr sum sumwgt min max var;
VAR interactions;
by cltype;
weight W33P0;
strata W33PSTR;
cluster W33PPSU;
RUN;

proc surveyreg;
weight W33P0;
strata W33PSTR;
cluster W33PPSU;
class cltype;
model care_safety=cltype;
estimate 'mean inclusive' intercept 1 cltype 1;
RUN;
QUIT;

proc surveymeans mean nobs std stderr sum sumwgt min max var;
VAR care_safety;
by cltype;
weight W33P0;
strata W33PSTR;
cluster W33PPSU;
RUN;
quit;

```

SAS™ Syntax to Conduct Group Comparisons of ECERS-R Derived Composites

```

PROC IMPORT OUT= WORK.crystal
      DATAFILE= "C:\Users\Restrict\Desktop\Crystal\ECLS-B\SPSS Dat
a Files\recoded_CISCOMPOSITES-7.3.13.sav"
      DBMS=SPSS REPLACE;

RUN;

data arnettcomposite;
set Crystal;
RUN;
QUIT;

proc sort;
by cltype;
RUN;
QUIT;

proc surveyreg;
weight W33P0;
strata W33PSTR;
cluster W33PPSU;
class cltype;
model careint=cltype;
estimate 'mean inclusive' intercept 1 cltype 1;

```

```

RUN;
QUIT;

proc surveymeans mean nobs stderr min max var;
VAR careint;
by cltype;
weight W33P0;
strata W33PSTR;
cluster W33PPSU;
RUN;

proc surveyreg;
weight W33P0;
strata W33PSTR;
cluster W33PPSU;
class cltype;
model negative=cltype;
estimate 'mean inclusive' intercept 1 cltype 1;
RUN;
QUIT;

proc surveymeans mean nobs stderr min max var;
VAR negative;
by cltype;
weight W33P0;
strata W33PSTR;
cluster W33PPSU;
RUN;

proc surveyreg;
weight W33P0;
strata W33PSTR;
cluster W33PPSU;
class cltype;
model positive=cltype;
estimate 'mean inclusive' intercept 1 cltype 1;
RUN;
QUIT;

proc surveymeans mean nobs stderr min max var;
VAR positive;
by cltype;
weight W33P0;
strata W33PSTR;
cluster W33PPSU;
RUN;
quit;

```

REFERENCE LIST

- Abbot-Shinn, M., & Sibley, A. (1992). *Assessment profile for early childhood programs: Research version*. Atlanta, GA: Quality Assist.
- Abt Associates Inc. (2006). *Observation training manual: OMLIT early childhood* [training manual]. Cambridge, MA: Author.
- Americans with Disabilities Amendments Act, 42 U.S.C.A. § 120101 (2008) (Amended from the American Disabilities Act of 1990)
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Arnett, J. (1986). *Caregivers in day-care centers: Does training matter?* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8705689).
- Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10, 541-552.
- Bailey, D. B., Clifford, R. M., & Harms, T. (1982). Comparison of preschool environments for handicapped and nonhandicapped children. *Topics in Early Childhood Special Education*, 2, 9-20.
doi:<http://dx.doi.org/10.1177/027112148200200106>
- Berkely, T. R., & Ludlow, B. L. (1989). Toward a reconceptualization of the developmental model. *Topics in Early Childhood Special Education*, 9, 51-66.
- Barrett, P. (2007), Structural equation modeling: Judging model fit. *Personality and Individual Differences*, 42, 815-24.
- Bentler, P. M. and Bonnet, D. C. (1980), Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588-606.
- Bollen, K. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Bredenkamp, S. (Ed.). (1987). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8*. Washington, DC: National Association for the Education of Young Children.
- Bredenkamp, S., & Copple, C. (Eds.) (1997). *Developmentally appropriate practice in early childhood programs* (Revised ed.). Washington, DC: National Association for the Education of Young Children.

- Bronfenbrenner, U. & Crouter, A. C. (1983). The evolution of environmental models in developmental research. In W. Kessen (Series Ed.) & P. H. Mussen (Vol. Ed.), *Handbook of child psychology: History, theory, and methods* (4th ed., Vol. 1, pp. 357-414), New York, NY: Wiley.
- Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental processes. In W. Damon (Series Ed.) & R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Theoretical models of human development* (5th ed., Vol. 1, pp. 993-1028). New York: Wiley.
- Bruder, M. B., & Brand, M. (1995). A comparison of two types of early intervention environments serving toddler-age children with disabilities. *Infant-Toddler Intervention: The Transdisciplinary Journal*, 5, 207-217.
- Bryant, D. M., Burchinal, M., Lau, L. B., & Sparling, J. J. (1994). Family and classroom correlates of head start children's developmental outcomes. *Early Childhood Research Quarterly*, 9, 289-304. doi:10.1016/0885-2006(94)90011-6
- Bryant, D. M., Burchinal., M., & Zaslow, M. (2011). Empirical approaches to strengthening the measurement quality: Issues in the development and use of quality measures in research and applied settings. In M. Zaslow, I. Martinez-Beck, K., Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings*. Blatimore, MD: Brookes.
- Burchinal, M. R., Cryer, D., Clifford, R., & Howes, C. (2002). Caregiver training and classroom quality in child care centers. *Applied Developmental Science*, 6, 2-11.
- Burchinal, M., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher-child interactions and instruction. *Applied Developmental Science*, 12, 140-153.
- 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*. Blatimore, MD: Brookes.
- Burchinal, M. R., Peisner-Feinberg, E., Bryant, D. M., & Clifford, R. (2000). Children's social and cognitive development and child-care quality: Testing for differential associations related to poverty, gender, or ethnicity. *Applied Developmental Science*, 4, 149-165. doi:10.1207/S1532480XADS0403_4
- Buyse, V., & Bailey, D. B. (1993). Behavioral and developmental outcomes in young children with disabilities in integrated and segregated settings: A review of comparative studies. *The Journal of Special Education*, 26, 434-461.
- Buyse, V., Wesley, P. W., Bryant, D., & Gardner, D. (1999). Quality of early childhood programs in inclusive and noninclusive settings. *Exceptional Children*, 65, 301-314.

- Byrne, B.M. (1998). *Structural equation modeling with LISREL, PRELIS and SIMPLIS: Basic concepts, applications and programming*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Carta, J. J., Schwartz, I. s., Atwater, J. B., & McConnell, S. R. (1991). Developmentally appropriate practice: Appraising its usefulness for young children with disabilities. *Topics in Early Childhood Special Education*, 11, 1-20.
- 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, 505-520.
- Cassidy, D. J., Hestenes, L. L., Hegde, A., Hestenes, S., & Mims, S. (2005). Measurement of quality in preschool child care classrooms: An exploratory and confirmatory factor analysis of the early childhood environment rating scale-revised. *Early Childhood Research Quarterly*, 20, 345-360.
- Ceglowski, D. (2004). How stake holder groups define quality in child care. *Early Childhood Education Journal*, 32, 101-111.
- Ceglowski, D., & Bacigalupa, C. (2002). Four perspectives on child care quality. *Early Childhood Education Journal*, 30, 87-92.
- Clarke-Stewart, K. A., Vandell, D. L., Burchinal, M., O'Brien, M., & McCartney, K. (2002). Do regulatable features of child-care homes affect children's development? *Early Childhood Research Quarterly*, 17, 52-86.
- Clawson, C., & Luze, G. (2008). Individual experiences of children with and without disabilities in early childhood settings. *Topics in Early Childhood Special Education*, 28, 132-147.
- Clifford, R. M., Barbarin, O., Chang, F., Early, D., Bryant, D., Howes, C., . . . Pianta, R. (2005). What is pre-kindergarten? Characteristics of public pre-kindergarten programs. *Applied Developmental Science*, 9, 126-143.
- Colwell, N., Gordon, R. A., Fujimoto, K., Kaestner, R., & Korenman, S. (2013). New evidence on the validity of the arnett caregiver interaction scale: Results from the early childhood longitudinal study-birth cohort. *Early Childhood Research Quarterly*, 28, 218-233. doi:10.1016/j.ecresq.2012.12.004
- Copple, C., & Bredekamp, S. (Eds.). (2009). *Developmentally appropriate practice in early childhood programs serving children from birth through age 8* (3rd ed.). Washington, DC: National Association for the Education of Young Children.
- Cryer, D. (1999). Defining and assessing early childhood program quality. *The ANNALS of the American Academy of Political and Social Science*, 563, 39-55. doi:10.1177/0002716299563001003

Cunningham, D. D. (2010). Relating preschool quality to children's literacy development. *Early Childhood Education Journal*, 37, 501-507.

Data Accountability Center. (2011). *Educational environment data tables* [Excel data tables]. Retrieved from https://www.idea-data.org/arc_toc13.asp#partbLRE

Divisions for Early Childhood's Task Force on Recommended Practices. (1993). *DEC recommended practices: Indicators of quality in programs for infants and young children with special needs and their families*. Reston, VA: Council for Exceptional Children.

Division for Early Childhood, National Association for the Education of Young Children. (2009). *Early childhood inclusion: A joint position statement of the Division for Early Childhood (DEC) and the National Association for the Education of Young Children (NAEYC)*. Chapel Hill: The University of North Carolina, FPG Child Development Institute.

Downer, J. T., Lopez, M. L., Grimm, K. J., Hamagami, A., Pianta, R. C., & Howes, C. (2012). Observations of teacher-child interactions in classrooms serving latinos and dual language learners: Applicability of the classroom assessment scoring system in diverse settings. *Early Childhood Research Quarterly*, 27, 21-32.

Dunn, L. (1993). Proximal and distal features of day care quality and children's development. *Early Childhood Research Quarterly*, 8, 167-192.

Early, D. M., Bryant, D. M., Pianta, R. C., Clifford, R. M., Burchinal, M. R., Ritchie, S., . . . Barbarin, O. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in pre-kindergarten? *Early Childhood Research Quarterly*, 21, 174-195.

Environment Rating Scale Institute. *Early childhood environment rating scale: Revised edition*. Retrieved from <http://www.ersi.info/ecers.html>

Gordon, R. A., Fujimoto, K., Kaestner, R., Korenman, S., & Abner, K. (2013). An assessment of the validity of the ECERS-R with implications for measures of child care quality and relations to child development. *Developmental Psychology*, 49, 146-160. doi: 10.1037/a0027899

Grisham-Brown, J., Cox, M., Gravil, M., & Missall, K. (2010). Differences in child care quality for children with and without disabilities. *Early Education and Development*, 21, 21-37.

Halle, T., Whittaker, J. V., & Anderson, R. (2010). *Quality in early childhood care and education settings: A compendium of measures* (2nd ed.). Washington, DC: Child Trends.

Harms, T., & Clifford, R. M. (1980). *Early childhood environment rating scale*. New York, NY: Teachers College Press.

- Harms, T., & Clifford, R. M. (1982). Assessing preschool environments with the early childhood environment rating scale. *Studies in Educational Evaluation*, 8, 261-269.
- Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early childhood environment rating scale* (Revised ed.). New York, NY: Teachers College Press.
- Harms, T., Clifford, R. M., & Cryer, D. (2005). *Early childhood environment rating scale* (Revised ed.). New York, NY: Teachers College Press.
- Harrist, A. W., Thompson, S. D., & Norris, D. J. (2007). Defining quality child care: Multiple stakeholder perspectives. *Early Education and Development*, 18, 305-336.
- Helburn, S. W. (1995). *Cost, quality and child outcomes in child care centers. Technical report, public report, and executive summary*. Denver, CO: Economics Department, University of Colorado at Denver.
- Hestenes, L. L., Cassidy, D. J., Shim, J., & Hegde, A. V. (2008). Quality in inclusive preschool classrooms. *Early Education and Development*, 19, 519-540.
- High/Scope. (2003). *Preschool program quality assessment*: (2nd ed.). Ypsilanti, MI: High/Scope Press.
- Horn, J. L., & McArdle, J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18, 114-125.
- Horn, J. L., McArdle, J., & Mason, R. (1983). When is invariance not invariant: A practical scientists look at the ethereal concept of factor invariance. *The Southern Psychologist*, 1, 179-188.
- Horowitz, F. D. (1987). *Exploring developmental theories: Toward a structural/behavioral model of development*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc.
- Howes, C., Phillips, D. A., & Whitebook, M. (1992). Thresholds or quality: Implications for the social development of children in center-based child care. *Child Development*, 63, 449-460.
- Hu, L. T. and Bentler, P. M. (1999), Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Hui, H. C., Triandis, H. C. (1985). Measurement in cross-cultural psychology: a review and comparison of strategies. *Journal of Cross-Cultural Psychology*, 16, 131-152.

Hyson, M., Hirsh-Pasek, K., & Rescorla, L. (1990). The classroom practices inventory: An observation instrument based on NAEYC's guidelines for developmentally appropriate practices for 4- and 5-year-old children. *Early Childhood Research Quarterly*, 5, 475-494.

Individuals with Disabilities Education Improvement Act of 2004, 20 U.S.C. § 1400 et seq. (2004)(reauthorization of the Individuals with Disabilities Act of 1990)

Institute of Education Sciences, National Center for Education Statistics. *Early childhood longitudinal program. Instruments and Assessments*. Retrieved from <http://nces.ed.gov/ecls/birthinstruments.asp>

Improving Head Start for School Readiness Act of 2007, U.S.C. § et. seq. (2007)

Institute of Education Sciences, National Center for Education Statistics. *Early childhood longitudinal program. Instruments and Assessments*. Retrieved from <http://nces.ed.gov/ecls/birthinstruments.asp>

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409-426.

Jöreskog, K. and Sörbom, D. (1993), *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago, IL: Scientific Software International Inc.

Knoche, L., Peterson, C. A., Edwards, C. P., & Jeon, H. (2006). Child care for children with and without disabilities: The provider, observer, and parent perspectives. *Early Childhood Research Quarterly*, 21, 93-109. doi: 10.1016/j.ecresq.2006.01.001

Lamorey, S., & Bricker, D. D. (1993). Integrated programs: Effects on young children and their parents. In C. A. Peck, S.L. Odom, & D. Bricker (Eds.), *Integrating young children with disabilities into community programs: Ecological perspectives on research and implementation*. Baltimore, MD: Brookes.

La Paro, K. M., Sexton, D., & Snyder, P. (1998). Program quality characteristics in segregated and inclusive early childhood settings. *Early Childhood Research Quarterly*, 13, 151-167.

La Paro, K. M., Thomason, A. C., Lower, J. K., Kintner-Duffy, V. L., & Cassidy, D. J. (2012). Examining the definition and measurement of quality in early childhood education: A review of studies using the ECERS-R from 2003 to 2010. *Early Childhood Research & Practice*, 14(1). Retrieved from http://ecrp.uiuc.edu/v14n1/laparo.html?goback=.gde_1807282_member_128314736

Laughlin, L. (2013). *Who's minding the kids? Child care arrangements: Spring 2011* (Report No. P70-135). Washington, DC: U.S. Census Bureau. Retrieved from <http://www.census.gov/prod/2013pubs/p70-135.pdf>

- Layzer, J. I., & Goodson, B. D. (2006). The “quality” of early care and education settings: Definitional and measurement issues. *Evaluation Review*, 30, 556-576. doi: 10.1177/0193841X06291524
- Layzer, J. I., Goodson, B. D., & Moss, M. (1993). *Observational studies of early childhood programs: Life in preschool final report*. Cambridge, MA: Abt Associates.
- Love, J., Ryer, P., & Faddis, B. (1992). *Caring environments-program quality in California's publicly funded child development programs. Final report on the legislatively mandated 1990-91 staff/child ratio study*. Portsmouth, NH: RMC Research Corporation.
- Love, J. M., Schochet, P. Z., & Meckstroth, A. L. (1996). *Are they in any real danger? What research does--and doesn't--tell us about child care quality and children's well-being*. Princeton, NJ: Mathematica Policy Research.
- 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, 113-127.
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., . . . Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79, 732-749.
- McLean, M., Hemmeter, M. L., & Snyder, P. (2014). *Essential elements for assessing infants and preschoolers with special needs*. Upper Saddle River, NJ: Pearson.
- Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, 58, 525-543
- Metz, A., & Bartley, L. (2012). Active implementation frameworks for program success: How to use implementation science to improve outcomes for children. *Zero to Three*, 32 (4), 11-18.
- Metz, A., Halle, T., Bartley, L., & Blasberg, A. (2013). The key components of successful implementation. In T. Halle, A. Metz, & I.M. Beck (Eds.), *Applying implementation science in early childhood programs and systems* (pp. 21-42). Baltimore: Brookes.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.
- Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate behavioral research*, 39, 479-515.

Mitchell, A. W. (2005). *Stair steps to quality: A guide for states and communities developing quality rating systems for early care and education*. Alexandria, VA: United Way of America, Success by 6.

Muthén, B. O. and Asparouhov (2002). Latent Variable Analysis With Categorical Outcomes: Multiple-Group And Growth Modeling In Mplus. *Mplus Web Notes*; No. 4 version 5. Retrieved from
<http://www.statmodel.com/download/webnotes/CatMGLong.pdf>

Muthén, L.K. & Muthén, B.O. (1998-2012). *Mplus User's Guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

National Association for the Education of Young Children. (2013). *NAEYC Accreditation: All criteria document*. Retrieved from:
<http://www.naeyc.org/files/academy/file/AllCriteriaDocument.pdf>

National Institute of Child Health and Human Development Early Child Care Research Network. (1998). Early child care and self-control, compliance, and problem behavior at twenty-four and thirty-six months. *Child Development*, 69, 1145-1170.

National Institute of Child Health and Human Development Early Child Care Research Network. (2000). The relation of child care to cognitive and language development. *Child Development*, 71, 960-980.

National Institute of Child Health and Human Development Early Child Care Research Network. (2002). Early child care and children's development prior to school entry: Results from the NICHD study of early child care. *American Educational Research Journal*, 39, 133-164.

National Institute of Child Health and Human Development Early Child Care Research Network. (2003). Does quality of child care affect child outcomes at age 4 ½? *Developmental Psychology*, 39, 451-469.

National Institute of Child Health and Human Development Early Child Care Research Network. (2006). Child-care effect sizes for the NICHD study of early child care and youth development. *American Psychologist*, 61, 99-116. doi: 10.1037/0003-066.X.61.2.99

National Institute for Early Education Research. (2007). *Preschool classroom mathematics inventory*. Rutgers, NJ: Author.

National Institute for Early Education Research. (2005). *The state of preschool*. Rutgers, NJ: Author.

National Institute of Health. (2012). *NICHD study of early child care and youth development (SECCYD)*. Retrieved from
<http://www.nichd.nih.gov/research/supported/Pages/seccyd.aspx>

Odom, S. L., & McEvoy, M. A. (1988). Integration of young children with handicaps and normally developing children. In S. L. Odom & M. B. Karnes (Eds.). *Early intervention for infants and children with handicaps: An empirical base*. Baltimore, MD: Brookes.

Odom, S.L., & McLean, M.E. (1996). Early intervention/early childhood special education: Recommended practices. Austin, TX: Pro-Ed.

Office of Planning, Research, & Evaluation. (1996-2010). *Early Head Start research and evaluation project (EHSRE): Project overview*. Retrieved from <http://www.acf.hhs.gov/programs/opre/research/project/early-head-start-research-and-evaluation-project-ehsre-1996-2010>

Office of Planning, Research, & Evaluation. (1997-2013). *Head Start family and child experiences survey (FACES), 1997-2013: Project overview*. Retrieved from <http://www.acf.hhs.gov/programs/opre/research/project/head-start-family-and-child-experiences-survey-faces-1997-2013>

Peisner-Feinberg, E. S., & Burchinal, M. R. (1997). Relations between preschool children's child-care experiences and concurrent development: The cost, quality, and outcomes study. *Merrill-Palmer Quarterly*, 43, 451-477.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., 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, 1534-1553.

Peterson, C. A., Wall, S., Raides, H. A., Kisker, E. E., Swanson, M. E., Jerald, J., . . .

Qiao, W. (2004). Early Head Start: Identifying and serving children with disabilities, *Topics in Early Childhood Special Education*, 24, 76-88.

Phillips, D. A., & Howes, C. (1987). Indicators of quality in child care: Review of research. In D. A. Phillips (Ed.), *Quality in child care: What does research tell us?* Washington, DC: National Association for the Education of Young Children.

Pianta, R. C., La Paro, K., & Hamre, B. (2008). *Classroom Assessment Scoring System—PreK [CLASS]*. Baltimore, MD: Brookes.

Perlman, M., Zellman, G. L., & Le, V. (2004). Examining the psychometric properties of the early childhood environment rating scale-revised (ECERS-R). *Early Childhood Research Quarterly*, 19, 398-412.

QRIS National Learning Network, BUILD Initiative. (August, 2013). *Current Status of QRIS in States*. Retrieved from <http://qrismetwork.org/sites/all/files/maps/QRIS%20Map,%20QRIS%20National%20Learning%20Network,%20www.qrisnetwork.org%20%5BRevised%20August%202013%5D.pdf>

- Ritchie, S., Howes, C., Kraft-Sayre, M., & Weiser, B. (2001). *Emerging academic snapshot*. Los Angeles, CA: University of California.
- Rock, D. A., Werts, C. E., & Flaugher, R. L. (1978). The use of analysis of covariance structures for comparing the psychometric properties of multiple variables across populations. *Multivariate Behavioral Research*, 13, 403-418.
- Ruopp, R., Travers, J., Glantz, F., & Coelen, C. (1979). *Children at the center: Final report of the National Day Care Study*. Cambridge, MA: Abt Associates.
- Rust, K. (1985). Variance estimation for complex estimation in sample surveys. *Journal of Official Statistics*, 1, 381-397.
- Rutter, M. (1987). Psychosocial resilience and protective mechanisms. *American Journal of Orthopsychiatry*, 57, 316-331.
- Rutter, M.,...Eaves, L. (1997). Integrating nature and nurture: Implications of person-environment correlations and interactions for developmental psychopathology. *Development and Psychopathology*, 9, 335-364.
- Sakai, L. M., Whitebook, M., Wishard, A., & Howes, C. (2003). Evaluating the early childhood environment rating scale (ECERS): Assessing differences between the first and revised edition. *Early Childhood Research Quarterly*, 18, 427-445.
- Sameroff, A. J., & Chandler, M. (1975). Reproductive risk and the continuum of caretaking casualty. In F. Horowitz, M. Hetherington, S. Scarr-Salapatek, & G. Siegel (Eds.), *Review of child development research* (Vol. 4, pp. 187-244). Chicago: University Chicago Press.
- Sandall, S., Hemmeter, M.L., Smith, B.J., & McLean, M.E. (Eds.).(2005). *DEC recommended practices in early intervention/ early childhood special education: A comprehensive guide for practical application*. Longmont, CO: Sopris West.
- Sandall, S., McLean, M.E., & Smith, B.J. (2000). DEC recommended practices in early intervention/early childhood special education. Longmont, CO: Sopris West.
- Scarr, S., Eisenberg, M., & Deater-Deckard, K. (1994). Measurement of quality in child care centers. *Early Childhood Research Quarterly*, 9, 131-151. doi: 10.1016/0885-2006(94)90002-7
- Shonkoff, J., & Phillips, D. A. (Eds.) (2000). *From neurons to neighborhoods*. Washington, DC: National Academy of Sciences.
- Smith, M., Dickinson, D., Sangeorge, A., & Anastasopoulos, L. (2002). *Early language and literacy classroom observation (ELLCO) toolkit, research edition*. Baltimore, MD: Brookes.

- Snow, C. E., & Van Hemel, S. B. (Eds.). (2008). *Early childhood assessment: Why, what, and how*. Washington, DC: National Academies Press.
- Snow, K., Derecho, A., Wheless, S., Lennon, J., Rosen, J., Rogers, J., . . . Einaudi, P. (2009). *Early childhood longitudinal study, birth cohort (ECLS-B): Kindergarten 2006 and 2007 data file user's manual*. (NCES 2010-010). Washington, DC: U. S. Dept. of Education, Institute of Education Sciences, National Center for Education Statistics.
- Snow, K., Thalji, L., Derecho, A., Wheless, S., Lennon, J., Kinsey, S., . . . Park, J. (2007). *Early childhood longitudinal study, birth cohort (ECLS-B): Preschool year data file user's manual (2005-2006)*. (NCES 2008-024). Washington, DC: U. S. Dept. of Education, Institute of Education Sciences, National Center for Education Statistics.
- Soukakou, E. P. (2012). Measuring quality in inclusive preschool classrooms: Development and validation of the inclusive classroom profile (ICP). *Early Childhood Research Quarterly*, 27, 478-488.
- Spiker, D., Hebbeler, K. M., & Barton, L. M. (2011). Measuring quality of ECE programs for children with disabilities. In M. Zaslow, I. Martinez-Beck, K., Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings*. Blatimore, MD: Brookes.
- Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78-90.
- Stypek, D., & Byler, P. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly*, 19, 375-397.
- Sylva, K., Siraj-Blatchford, I., & Taggart, B. (2003). *Assessing quality in the early years: Early childhood environment rating scale-Extension (ECERS-E): Four curricular subscales*. Stoke-on Trent, Staffordshire, England: Trentham Books.
- Thompson, B. T. (2002). What future quantitative social science research could look like: Confidence intervals for effect sizes. *Educational Researcher*, 31, 24-31.
- Thompson, B. T. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools*, 44, 423-432. doi: 10.1002/pits.20234
- 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. Dept. 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*. Washington, DC: Child Trends

- Tout, K., Zaslow, M., Halle, T., & Forry, N. (2009). ACF-OPRE Issue Brief. *Issues for the next decade of quality rating and improvement systems*. Washington, DC: Office of Planning, Research and Evaluation, U.S. Dept. of Health and Human Services.
- U.S. Dept. of Commerce, United States Census Bureau. (2013). *About child care*. Retrieved from <http://www.census.gov/hhes/childcare/about/index.html>
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4-70. doi: 10.1177/109442810031002
- Vandell, D. L., & Wolfe, B. (2000). *Child care quality: Does it matter and does it need to be improved?* (Special Report No. 78). Madison, WI: University of Wisconsin-Madison, Institute for Research on Poverty.
- Vogel, C. A., Xue, Y., Moiduddin, E. M., Kisker, E. E., & Carlson, B. L. (2010). *Early Head Start children in grade 5: Long-term follow-up of the Early Head Start Research and Evaluation Study Sample* (OPRE Report No. 2011-8). Washington, DC: Office of Planning, Researcher, and Evaluation, Administration for Children and Families, US Dept. of Health and Human Services. Retrieved from <http://www.acf.hhs.gov/sites/default/files/opre/grade5.pdf>
- Votruba-Drzal, E., Levine Coley, R., & Lindsaychase-Lansdale, P. (2004). Child care and low-income children's development: Direct and moderated effects. *Child Development*, 75, 296-312.
- Whitebook, M., Howes, C., & Phillips, D. (1989). *Who cares? Child care teachers and the quality of care in America. Final report of the Child Care Staffing Study*. (Report No. PS 019045). Oakland, CA: Child Care Employee Project.
- Wilkinson, L., & Task force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594-604.
- Wolery, M., & Wilbers, J. S. (Eds.). (1994). *Including children with special needs in early childhood programs*. Washington, DC: National Association for the Education of Young Children.
- 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. Dept. of Health and Human Services.
- Zellman, G. L., Perlman, M., Le, V., & 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. (MG-650-QUE). Santa Monica, CA: RAND Corporation.

BIOGRAPHICAL SKETCH

Crystal Bishop earned her doctoral degree from the University of Florida in Fall 2013. Her major area of study was special education, with a minor in research evaluation and methodology and concentrations in early childhood studies and special education policy. Crystal's research interests include (a) the ways in which early childhood and special education policies are translated and enacted as early childhood practices; (b) measurement of the implementation of evidence-based practices in early childhood settings; and (c) professional development for teachers and administrators supporting young children with and without disabilities in inclusive settings. In 2013, Crystal received the J. David Sexton Doctoral Student Award from the Council for Exceptional Children—Division of Early Childhood (DEC). The award is given to a DEC member and doctoral-level student who has made significant contributions to young children with special needs and their families through their efforts in research, higher education, publications, policy, and information dissemination.

Prior to enrolling at the University of Florida, Crystal completed a bachelor's degree in Zoology and Physiology at the University of Wyoming. She received a master's degree in Human Development Counseling from the Vanderbilt University. During her master's program, Crystal was awarded a trainee grant in the Leadership Education in Developmental and Related Disabilities (LEND) program. The LEND program is designed to provide high-quality interdisciplinary training to professionals from diverse disciplines to improve the health of infants, children, and adolescents with disabilities.

Following the completion of her doctoral degree, Crystal was awarded a post-doctoral fellowship at the University of Florida Center for Excellence in Early Childhood

Studies. The Center is a campus-wide interdisciplinary center focused on the science of early childhood development and learning.