EXPLORING THE RELATIONSHIP BETWEEN GLOBAL QUALITY AND CHILD ENGAGEMENT IN TODDLER CHILD CARE CLASSROOMS

by

Alison Hooper

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Human Development and Family Studies

Fall 2013

© 2013 Alison Hooper All Rights Reserved

EXPLORING THE RELATIONSHIP BETWEEN GLOBAL QUALITY AND CHILD ENGAGEMENT IN TODDLER CHILD CARE CLASSROOMS

by

Alison Hooper

Approved:

Rena Hallam, Ph.D. Professor in charge of thesis on behalf of the Advisory Committee

Approved:

Susan J. Hall, Ph.D. Interim Chair of the Department of Human Development and Family Studies

Approved:

Lynn R. Okagaki, Ph.D. Dean of the College of Education and Human Development

Approved:

James G. Richards, Ph.D. Vice Provost for Graduate and Professional Education

ACKNOWLEDGMENTS

I would like to thank the many people who provided help and encouragement during the process of writing this thesis. Many thanks to Dr. Rena Hallam for her time spent discussing ideas, providing feedback, and reading drafts, as well as her willingness to let me partner with Delaware Stars to complete the study. To my committee members, Dr. Myae Han and Dr. Cynthia Paris, thank you for your support, ideas, and feedback. Kaitlin Bargreen, thanks for your invaluable assistance in adapting the engagement measure and collecting data and for sharing your knowledge and experience. To the assessment team of Delaware Stars, your support for this project and your willingness to let me join you on assessments made this research possible. I always enjoy the chance to work with you. Thanks to my HDFS graduate student friends, especially Kelley Perkins, who was always willing to brainstorm and discuss ideas. Lastly, thanks to my family and especially my husband for his patience with the many hours spent on this project and for his encouragement.

TABLE OF CONTENTS

LIST	OF TABLES	vi
LIST (OF FIGURES	vii
ABST	RACT	viii
Chant		
Chapt	er	
1	INTRODUCTION	1
2	THEORETICAL FRAMEWORK AND LITERATURE REVIEW	3
	Process-Person-Context-Time Model	3
	Engagement as a Proximal Process	5
	Child Care for Infants and Toddlers	6
	Defining Child Care Quality	8
	Measuring Child Care Quality	10
	Child Care Quality from a Bottom-Up Perspective	14
	Children's Engagement in Child Care	18
	Measuring Engagement in Child Care Classrooms	
3	METHODS	
	Contant of the Study	22
	Context of the Study	
	Sample	
	Measures	
	Procedures	
4	RESULTS	34
•		
	The Relationship between Engagement and Global Quality	
	Differences in Engagement between Classrooms by Global Quality	
	Differences in Engagement and Global Quality by Star Level	
	Engagement by Classroom Context	
5	DISCUSSION	
	Engagement and Global Quality	
	Differences by Star Level	
	Engagement and Classroom Context	

Limitations	47
Implications and Future Directions	
RENCES	50
dix	
ENGAGEMENT CODING SHEET IRB LETTER	64 65
	Limitations Implications and Future Directions RENCES dix ENGAGEMENT CODING SHEET IRB LETTER

LIST OF TABLES

Table 1	Teacher Demographic Data	. 25
Table 2	Program and Classroom Demographic Data	. 26
Table 3	Definitions and Common Examples of Contexts	. 29
Table 4	Frequency of Observations by Context	. 33
Table 5	Descriptive Statistics of Observational Data by Classroom	. 34
Table 6	Pearson Correlations between ITERS-R Subscales and Overall Engagement	. 35
Table 7	Differences in Engagement by Mid-to-High and Low ITERS-R Scores	. 37
Table 8	Mean Engagement and ITERS-R Scores by Star Level	. 38
Table 9	Descriptive Statistics of Engagement Contexts by Classroom	. 40
Table 10	Descriptive Statistics of Engagement Contexts across All Observation Intervals	. 40
Table 11	Correlations between Engagement by Context and ITERS-R Composite Score	.41

LIST OF FIGURES

Figure 1	Scatter Plot of Relationship between ITERS-R and Child	
	Engagement	36

ABSTRACT

Toddlers' engagement with their social and physical environment is an important aspect of their experience in early care and education programs. The purpose of this research study was to examine how global quality relates to children's engagement in toddler child care classrooms. Additionally, this study explored how toddlers' group engagement levels vary across classroom contexts, including free play, group activities, meals, transitions, and personal care routines. Thirty toddler child care classrooms participating in Delaware's Quality Rating and Improvement System (QRIS) were observed using two observational measures. Results indicate that a strong positive relationship exists between engagement and global quality. Global quality and engagement varied significantly between Star Level 2 and Star Level 4 in the QRIS. Child engagement varied significantly by classroom context, with the highest levels of engagement documented during mealtime and free play and the lowest levels of engagement documented during transitions.

Chapter 1

INTRODUCTION

Child care quality is increasingly an important area of research. Providing quality child care for very young children found its way into the research agenda of the new millennium (Melhuish, 2001) and is increasingly on the public radar due to recent national initiatives like Race to the Top Early Learning Challenge. With the expansion of Quality Rating and Improvement Systems (QRIS), the high stakes for child care centers to earn the distinction "high quality," and the desire to give children positive early experiences that will prepare them for school success, it is crucial to consider how quality is being defined and how to best measure quality. However, quality remains difficult to define, due to differing ideas about what is most important from different stakeholder groups (Ceglowski, 2004; Ceglowski & Bacigalupa, 2002; Harrist, Thompson, & Norris, 2007).

In the midst of defining and measuring child care quality, it is important to pause and consider if the current definitions and measures in the field capture what matters most for young children. There is increasing research support for changing or supplementing the current predominant approaches to defining and measuring quality so that they more directly relate to young children's experiences (Hallam, Fouts, Bargeen, & Caudle, 2009; Raspa, McWilliam, & Ridley, 2001; Tonyan & Howes, 2003). Although there is some effort to enhance quality measures in preschool classrooms (Pianta, La Paro, & Hamre, 2008; Sylva, Siraj-Blatchford, & Taggart, 2003), there are fewer options available for infant and toddler classrooms, where research has shown that the quality of a child's environment may greatly impact later child outcomes (Ackerman & Barnett, 2009; Burchinal, Nabors, Bryan, & Roberts, 1996). One promising strategy for better capturing children's experiences is examining children's engagement as an indicator of quality (Raspa et al., 2001; Ridley, McWilliam, & Oates, 2000). This study will focus on using engagement as an indicator for classroom quality in toddler classrooms.

Chapter 2

THEORETICAL FRAMEWORK AND LITERATURE REVIEW

Bioecological theory (Bronfenbrenner, 1998) is a helpful lens through which to consider the relation between engagement and child care quality. According to bioecological theory, children's development occurs within the context of the complex systems in their environment (1979). Early ecological theory focuses more on the different environments in which interactions occur than on the interactions themselves, considering four nested systems which affect children's development (Bronfenbrenner & Morris, 1998). The microsystem considers the relations between a person and their immediate setting. The next system, the mesosystem, comprises the interactions between settings. The exosystem consists of influences on the child that do not directly include members of the microsystem. Next, the macrosystem includes the overarching social, political, and cultural influences which affect the other systems. Another level, the chronosystem, was added later, which frames the dynamics of the other systems in historical context (Bronfenbrenner, 1992).

Process-Person-Context-Time Model

Bronfenbrenner continued to refine and add to the ecological model of development, leading to the bioecological model, which has four main components: process, person, context, and time. Proximal processes are at the center of the Process-Person-Context-Time (PPCT) model. Proximal processes are the specific forms of interaction that occur between the individual and his or her environment that facilitate development (Bronfenbrenner, 1999). They can occur between a child and people, objects, or symbols in the immediate environment. According to bioecological theory, developmental outcomes are a result of the interaction of proximal processes and an individual's characteristics (Bronfenbrenner & Morris, 1998). Two propositions define the structure of the bioecological model. The first proposition states:

In order to develop—intellectually, emotionally, socially, and morally—a human being, whether child or adult, requires the same thing: active participation in progressively more complex, reciprocal interaction with persons, objects, and symbols in the individual's immediate environment. To be effective, the interaction must occur on a fairly regular basis over extended periods of time. Such enduring forms of interaction in the immediate environment are referred to as proximal processes. Proximal processes are posited as the primary engines of development. (Bronfenbrenner & Morris, 1998, p. 996)

The second proposition highlights the other components of the PPCT model:

Proximal processes cannot structure, steer or sustain themselves. Their form, power, content, and direction vary systematically as a joint function of the characteristics of the **developing person** and of the **environment**—both immediate and more remote—in which the processes are taking place; the **time** through the life course and the historical period during which the

person has lived; and the nature of the **developmental outcome** under consideration. (Bronfenbrenner & Morris, 1998, p. 996)

Proximal processes are central to studying children's engagement in the child care classroom setting. Proximal processes occur when children interact with teachers, peers, objects, and ideas. Person, context, and time are also part of the PPCT model. Person represents the child's characteristics when they enter the classroom environment. These characteristics are influenced by prior experiences and the child's risk and protective factors. Research has shown that child care can be a risk factor or a protective factor for children and families at risk depending on the quality of the care (Davies, 2004).

Context and time in PPCT incorporate the original system levels in Bronfenbrenner's ecological theory. Each of these systems affect a young child attending child care. One of a child's microsystems is the immediate family, but for children attending child care, the classroom is another microsystem, and the microsystem of most interest for this study. Classroom quality and child engagement are both microsystem-level variables (Odom et al., 2004). Although this study considers microsystem-level variables, it is clearly enmeshed in all of the systems that influence a child's development, including state and national policies around QRIS.

Engagement as a Proximal Process

Central to this study are the proximal processes that occur within a young child attending child care's microsystem, the classroom. For this study engagement will be

used to operationalize the measurement of one aspect of proximal processes. Engagement is defined as the amount of time children spend interacting with the environment in a developmentally appropriate and contextually appropriate manner (McWilliam & Bailey, 1995). Often research of children's experiences in child care focuses more narrowly on children's interactions with teachers or peers (Girolametto, & Weitzman, 2002; Hamre & Pianta, 2005; Thomason & La Paro, 2009), but engagement considers multiple components of proximal processes at once. Because proximal processes drive a child's development, anything that interferes with the occurrence of effective proximal processes can be harmful to the developing child (Bronfenbrenner & Evans, 2000). Low classroom quality, including infrequent responsive interactions with caregivers, lack of appropriate materials, and developmentally inappropriate classroom activities, may interfere with the occurrence of proximal processes. Therefore, it is hypothesized that lower-quality classrooms would have lower levels of child engagement.

Child Care for Infants and Toddlers

According to the Early Childhood Longitudinal Study, Birth Cohort, over half of children born in 2001 spent time in regular nonparental care at nine months (Flanagan & West, 2004), and this number continues to rise (Ackerman & Barnett, 2009; Kreader, Ferguson, & Lawrence, 2005a). Infants and toddlers in center-based settings are likely to attend care 31 to 40 hours per week (Ackerman & Barnett, 2009; Flanagan & West, 2004). Because many very young children are spending significant time in non-parental care, it is important to consider the quality and effects of this care.

High-quality care is especially important for infants and toddlers. Recent brain research shows that the early years are a time of great developmental opportunity (National Scientific Council on the Developing Child, 2004; Shonkoff & Phillips, 2000), but these opportunities can be missed if a child's environment does not provide the basic level of stimulation and nurturing that the brain needs (Knitzer & Lefkowitz, 2006). Unfortunately, infants and toddlers currently have the least access to highquality care (Ackerman & Barnett, 2009; Kreader, Ferguson, & Lawrence, 2005b). The 1995 Cost, Quality, and Child Outcomes Study (Helburn, 1995) found that infant and toddler care is of the poorest quality in the United States. In this study, 8% of infant and toddler classrooms were providing good quality care, as compared to 24% of preschool classrooms. 51% of infant and toddler classrooms were providing medium or mediocre care, and 40% were providing care of poor quality, while only 10% of preschool classrooms were rated poor-quality.

Another national study completed by the National Institute of Child Health and Human Development (NICHD) measured quality of infant and toddler care. It arose from the concern that attending child care may be harmful to infants' development, specifically to forming a secure attachment with their primary caregiver (Belsky, 1988). The NICHD Study of Early Child Care found that 10% of centers serving infants and toddlers were rated poor-quality, compared to only 4% of centers serving older children (NICHD, 2002). While the NICHD study found no direct effects of the quality or quantity of child care on attachment security, the study found that poor quality care combined with low maternal sensitivity was associated with an increased risk for insecure attachment (NICHD, 1997). This study and others suggest that lowquality child care can be a risk factor for infants and toddlers, while high-quality care

can be a protective factor (Caughy, DiPietro, & Strobino, 1994; Egeland & Hiester, 1995; NICHD, 1997; Melhuish, 2001). While a universal definition of quality is lacking, there are a number of important practices that have consistently been found to be present in high-quality settings for infants and toddlers. These include responsive interactions with caregivers (Girolametto & Weitzman, 2002; NICHD, 2002), safe and healthy environments (Cryer, 1999; Phillips et al., 2000), and time spent engaged with peers, adults, and materials (McWilliam et al., 1985; Ridley et al., 2000).

Defining Child Care Quality

Research has continued to demonstrate associations between child care quality and children's academic and social outcomes (Belsky et al., 2007; Cryer, 1999; Helburn, 1995; Howes et al., 2008; Mashburn et al., 2008). Because of these findings and because of the increasing number of children who are spending time in centerbased care (Flanagan & West, 2004), national and state policies have focused on improving the quality of child care for young children. While quality continues to be a major research focus in early care and education, challenges remain in defining and measuring this construct.

Early research identified proximal and distal features of classrooms that promote young children's development in different domains (Dunn, 1993). Proximal features, which describe children's actual experiences, include curriculum and classroom interactions, while distal features, which describe experiences potentially available to children, include program and state policies and other structural variables. This early study of child care quality found that proximal and distal features of quality were equally effective in predicting children's development. Definitions of child care quality have historically included multiple proximal and distal features of classrooms,

but there are many perspectives about which proximal and distal features are most important. This has led to definitions of quality that tend to be broad and nonspecific (La Paro, Thomason, Lower, Kinter-Duffy, & Cassidy, 2012; Layzer & Goodson, 2006). Additionally, other important aspects of early childhood environments are often diminished or left out completely in overgeneralized definitions of quality. Because quality is multidimensional, it is important to consider that general definitions and single global measures may not be able to adequately capture all important and meaningful aspects of quality.

Researchers typically refer to two main types of child care quality: process quality and structural quality. Process quality includes the aspects of the child care setting that children directly experience. This includes the child's interactions with adults, peers, and materials. Structural quality includes the factors that create the framework that allows these processes to occur (Cryer, 1999). Some examples of structural quality include a teacher's level of education, the teacher to child ratio of the classroom, and group size (Vandell & Wolfe, 2000). Process features more directly measure children's experiences in child care, but structural factors are simpler and less expensive to measure and easier to regulate (Kreader et al., 2005b). Therefore, researchers have tried to identify relationships between structural and process quality (Cryer, 1999). Research has shown that structural quality and process quality are related, but this relationship appears to be complex (Layzer & Goodson, 2006; Phillips et al., 2000; Vandell & Wolfe, 2000). No one structural variable strongly accounts for variations in process quality, so it is important to consider many aspects of structural quality together to try to improve process quality (Cryer, 1999). Some research has

identified relationships between specific structural variables and process variables, but these have yet to be replicated consistently in other studies.

QRIS are one strategy for increasing child care quality, as well as helping families identify and select high-quality care for their children. As of 2010, 26 states had statewide QRIS and all but four states were currently using, exploring, or piloting at QRIS either statewide or in specific regions (NACCRRA, 2009; Shulman, Matthews, Blank, & Hannah, 2012; Tout et. al, 2010). QRIS are a tool to improve families' access to high-quality care through rating individual child care programs and offering incentives and assistance to increase centers' ratings. QRIS address five basic elements: quality standards, accountability, program support, parent education, and financial incentives (Mitchell, 2005; NACCRRA, 2009). One of the primary tasks facing state QRIS is selecting an appropriate and effective measure of quality that can be implemented efficiently.

Measuring Child Care Quality

The lack of a unified definition of quality along with challenges around measuring quality from different stakeholder perspectives has led of a widespread reliance on the Environment Rating Scales (ERS), the most widely used measures of quality to date (La Paro et al., 2012; NACCRAA, 2009; Tout et al., 2010; Tout, Zaslow, Halle, & Forry, 2009). These include the Early Childhood Environment Rating Scale–Revised (ECERS-R; Harms, Clifford, & Cryer, 1998), the Infant/Toddler Environment Rating Scale–Revised (ITERS-R; Harms, Cryer, & Clifford, 2003), the Family Child Care Environment Rating Scale–Revised (Harms, Cryer, & Clifford, 2007), and the School-Age Care Environment Rating Scale (Harms, Jacobs, & White, 2006). In most large-scale studies measuring child care quality for infants and toddlers, researchers use the Infant/Toddler Environment Rating Scale (Harms, Cryer, & Clifford, 1990) or ITERS-R to measure process quality (Burchinal et al., 1996; Helburn, 1995; Howes, Phillips, & Whitebrook, 1992; Scarr, Eisenberg, & Deater-Deckard, 1994).

The ERS were originally designed as self-assessment tools for programs to use to identify strengths and weaknesses (Harms et al., 1998), but they are currently used widely in research studies and QRIS to assess programs' quality and inform policy decisions (La Paro et al., 2012; NACCRAA, 2009; Tout et al., 2009). As of 2009, 23 of 26 states implementing QRIS use the ITERS-R to measure quality (Tout et al., 2010). The ITERS-R contains 39 items organized in seven subscales measuring different areas of quality and was designed to measure quality as experienced by all children in a group (Cryer, 1999). Classrooms are given a score from one (inadequate) to seven (excellent) on each item based primarily on classroom observation with some staff interview, and item scores are averaged to get an overall quality score. Due to its widespread use, some stakeholders consider the ERS to be synonymous with quality (La Paro et al., 2012). However, La Paro et al. (2012) found that even research studies using the ERS do not use a consistent definition of quality.

With the ITERS-R being the primary tool for assessing process quality in infant and toddler classrooms and for determining a center's quality rating, which is frequently tied to funding, it is important to consider whether the ITERS-R is truly measuring classroom quality adequately. Because the ERS is widely used in QRIS as the primary evaluation tool, the individual items of the ERS have become a focus for quality improvement efforts as programs try to increase their ERS scores to earn higher ratings and incentives in the QRIS.

One widespread criticism of the ITERS-R is that it does not include all of the multiple dimensions of interactions that toddlers experience. Of the seven subscales of the ITERS-R, only two, Listening and Talking and Interaction, address the "how" of teaching behaviors (Thomason & La Paro, 2009). Additionally, it appears that the ITERS-R does not truly measure distinct seven distinct areas of quality, which is what it was designed to do, although there is not currently consensus on how many factors of quality it actually measures (Bisceglia, Perlman, Schaack, & Jenkins, 2009; Hestenes, Cassidy, Hegde, & Lower, 2007). Bisceglia et al. (2009) completed a factor analysis on the tool excluding the Parents and Staff subscale and found that all the indicators in the six subscales loaded on a single factor. Therefore, they determined that the instrument does not measure separate areas of quality as it was designed to do.

Hestenes et al. (2007) also completed a factor analysis of the ITERS-R and found four distinct dimensions of quality, which they identified as Materials/Activities, Language/Interactions, Safety/Organization, and Parents/Staff. Overall, research on the ITERS-R and ECERS-R has determined that the tools do not measure process quality as much as originally proposed, which means for the many states using the ERS as their primary quality measure tool, process quality is being vastly under-measured and underweighted in QRIS (Cassidy, Hestenes, Hegde, Hestenes, & Mims, 2005). Bisceglia et al. (2009) also identified that the ITERS-R may not be appropriate for use in high-stakes settings. Because the subscales are related, making a small change in the environment may make the center's quality appear higher than it really is. Providers are likely to make changes that will help them receive a higher score, but the items they change may not be reliable measures of quality. Since the ITERS-R was designed to measure the experiences of all children in

a group, it is also possible that a classroom earning a high composite score may not be meeting the needs of each of the individual children in the group.

The authors suggest that the ITERS-R has high reliability and validity (Environment Rating Scales Institute, n.d.; Harms et al., 2003), although this has been challenged by others using the measure in research and practice (Layzer & Goodson, 2006, Bisceglia et al., 2009; Hestenes et al., 2007). One reason the reliability and validity have been challenged is that the points on the scale, ranging from 1 to 7, may be not equally distant from each other (Layzer & Goodson, 2006). Another reason is the subscales may not measure distinct aspects of quality (Bisceglia et al., 2009; Hestenes et al., 2007). The authors of the ITERS-R cite Cohen's Kappa for the scale as .58 and for the scale without the Parents and Staff subscale as .55. They found a high level of internal consistency, with a Cronbach's alpha of .93, although the Space and Furnishings and Personal Care subscales both have Cronbach's alphas of below .6, which is below the level generally considered acceptable (Environment Rating Scales Institute, n.d.).

There is a growing discussion among QRIS leaders, researchers, and policy makers about the need for new measurement tools that better capture the multidimensionality of quality and children's experiences related to child outcomes and that are appropriate for high-stakes use (Bisceglia et al., 2009; Hallam et al., 2009; Thomason & La Paro, 2009; Tout et al., 2009), as well as the benefits of using multiple measure to more accurately capture quality (Denny, Hallam, & Homer, 2012; Dickinson, 2002). This is especially important for toddlers. Because of very young children's dependency on adults and their unique developmental characteristics, it is even more imperative than with other age groups to use measures of quality that

consider varying types of teacher behaviors and teacher-child interactions (Thomason & La Paro, 2009). Some states' QRIS, recognizing the need for better measures of process quality and the benefits of incorporating multiple measures, especially in high-stakes contexts, are currently supplementing the ECERS-R with other tools that more specifically focus on process quality (Pianta et al., 2008; Sylva et al., 2003; Tout et al., 2009), although this is being done much less with infants and toddlers.

Child Care Quality from a Bottom-Up Perspective

One reason child care quality remains difficult to define is that different stakeholder groups continue to emphasize different dimensions of quality (Ceglowski, 2004). Lilian Katz (1993) identified four perspectives on quality. The top-down perspective, which she identified as the most prevalent, consists of researchers' and policymakers' ideas about what quality in child care looks like. All major studies of child care quality have included the top-down perspective (Ceglowski & Bacigalupa, 2002), and the widely available measures to assess child care quality mostly contain characteristics selected by "experts" who are typically not directly involved in or impacted by child care quality (Melhuish, 2001). Although this perspective has some predictive power, it does not capture an individual child's experience (Katz, 1994). The other perspectives, which have not yet been as deeply investigated in research, include the outside-in perspective of parents, the inside-out perspective of caregivers and directors, and the bottom-up perspective of children.

Although the top-down approach to defining and measuring quality is still the most prevalent (Ceglowski & Bacigalupa, 2002; Harrist et al., 2007; Melhuish, 2001), the inside-out and outside-in approaches have received more research and public attention recently. The bottom-up approach to defining and measuring quality still

remains the least explored. This perspective considers what it is actually like to be a child in a specific program or classroom (Katz, 1994). Even though children are the group most affected by variations in child care quality, almost no one has considered their perspective (Ceglowski, 2004; Harrist et al., 2007). Katz (1993) was the first to publicly advocate for identifying quality from the child's perspective and the first to name this the bottom-up approach. She urged researchers to consider the daily quality of life each child was experiencing in child care and to ask themselves what it feels like to be a child in the environment in order to make meaningful judgments about a program's quality. Her rationale was that while the variables that researchers typically consider have been able to somewhat predict child outcomes, a high-quality environment as identified by the child may be more predictive of positive outcomes. According to Katz (1994), children have a real need to be deeply respected and intellectually engaged, ideas that are not often considered when measuring quality.

Quality from a bottom-up perspective is not synonymous with process quality in current research, but because process quality refers to what children directly experience (Cryer, 1999), they are closely related. When many researchers reference process quality, they are referring to global process quality (Melhuish, 2001), but the bottom-up perspective considers the view of the individual child. Quality from a bottom-up perspective would include factors such as comfort, a child's level of acceptance by the caregiver and peers, the level of engagement in meaningful activities he or she experiences, and the child's overall satisfaction with his or her experience (Ceglowski & Bacigalupa, 2002; Katz, 1993). It could be possible to earn a high quality score from the top-down perspective, for example having low group size and highly educated caregivers in a renovated facility, but earn a low quality score

from the bottom-up perspective, perhaps because caregivers do not regularly interact in meaningful ways with individual children. This illustrates the need for considering multiple perspectives, especially the bottom-up perspective, when defining and measuring quality. Research on child care quality has generally failed to illuminate children's daily lives in the child care setting (Ceglowski, 2004; Harrist et al., 2007; Katz, 1994; Raspa et al., 2001).

Introducing the bottom-up perspective into the child care quality conversation creates a number of complications. First, researchers must determine how to accurately and ethically capture children's perspectives. Katz (1993) suggested that highly trained evaluators use direct observation and extensive inference and that interviewing children directly about child care quality is unethical because it places children in a position where they may be asked to criticize their caregiver. Even so, some researchers have attempted to ask children directly about their experiences with and perceptions of child care (Ceglowski & Bacigalupa, 2007; Wiltz & Klein, 2001). These few studies where researchers have interviewed children directly have typically not included infants and toddlers because of their developmental levels. Although it is more difficult to include these very young children's voices in research, it is important. Figuring out what quality care looks like for this population may be a key to raising quality.

In addition to directly interviewing children, some researchers have used other methods to gain a bottom-up perspective on child care quality, one of which is time sampling observations of individual children (Hallam et al., 2009; Melhuish, 2001; Raspa et al., 2001; Tonyan & Howes, 2003) or of individual caregivers (Vandell & Wolfe, 2000). According to Melhuish (2001), this type of observation is more

effective than global quality measures like the ITERS-R. Perhaps the largest-scale use of a time sampling measure in child care research was in the NICHD Study of Early Child Care (NICHD, 1996), which used a measure called the Observation Record of the Caregiving Environment (ORCE) in infant/toddler and preschool classrooms. Through observations of individual children, the ORCE provided ratings of positive caregiving.

Even though a center may earn a high score on the ITERS-R, individual children's experiences in that center may vary widely. For example, Hallam et al. (2009) completed focal child observations in a toddler classroom of a top-rated center in the state's QRIS that had earned moderate ERS scores and found very low levels of interactions with toddlers. Although the classroom had earned lower scores on the two ITERS-R subscales addressing interactions, which are important to children's experiences and to later outcomes (Girolametto & Weitzman, 2002; NICHD, 2002), these lower scores were obscured by high scores in subscales that measure primarily structural quality and less directly capture children's daily experiences. These findings and others illustrate the need to consider other observational tools in addition to the ERS when measuring quality and reinforce that observing individual children in the context of the environment is a crucial part of measuring quality as it is experienced by the child (Hallam et al., 2009; Melhuish, 2001; NICHD, 1996; Thomason & La Paro, 2009).

Observing individual children, while giving a better view of an individual's experience, has limitations as well. Doing focal child observations is much more timeconsuming and expensive than measuring quality globally. Also, the results may not be generalizable to other children in the same classroom (Melhuish, 2001). Therefore,

it is important to examine other methods that may be able to capture children's daily experiences and be appropriate and practical for widespread use and in high-stakes contexts. One method that may be both effective and efficient is the use of environmental scans. An environmental scan uses time sampling, similar to focal child measures, but it considers the whole group of children at one time (Raspa et al., 2001; Ridley et al., 2000). While this approach is not able to capture the depth of information that can be obtained through focal child observations, it is a valuable tool to gather information about the whole group in one observation.

Children's Engagement in Child Care

Observing children's engagement is one way to measure the effects and quality of the child care environment (McWilliam et al., 1985). One benefit of measuring engagement is that it provides a more holistic picture of a child's experiences with one measure. Other measures that have focused primarily on process quality have tended to focus more narrowly on play or social behaviors (Raspa et al., 2001). Engagement is an important construct to examine because it helps to reveal how quality affects a child's daily experience, which is necessary before it is possible to understand how quality affects child outcomes (Ridley et al., 2000). Unlike global classroom quality, engagement focuses directly on children's experiences and behavior and can provide additional information that is not available through measuring only environmental quality.

While measures of global quality describe the context within which children's engagement occurs, it is necessary to look directly at children's engagement to understand what children are actually doing within the classroom context. Another important reason to consider child engagement in addition to global quality is that

measures of global quality may not adequately capture classroom dimensions that are most closely linked to child outcomes (Chien et al., 2010). Research studies show that children's positive engagement in classroom activities and routines contributes to later school achievement (Fredricks, Blumenfeld, & Paris, 2004; Greenwood, Horton, & Utley, 2002). It has also been found that children who spend more time engaged with their environment are more likely to develop positive behavior competencies (Ridley et al., 2000).

Previous research on engagement has primarily measured children's engagement levels across different classroom contexts and have found that context is an important variable in children's engagement (Booren, Downer, & Vitiello, 2012; Chien et. al, 2010; Downer, Rimm-Kaufman, & Pianta, 2007; Fuligni, Howes, Huang, Hong, & Lara-Cinisomo, 2012; Tonyan & Howes, 2003). In a study by Tonyan and Howes (2003) that examined how preschool children divided their time among specific activities and interactions using cluster analysis, children were more likely to be engaged in highly enriching activities in classrooms with higher global quality. Additionally, they found that the largest group of children spent most of their time in the least enriching activity clusters, gross motor and non-play, regardless of the classroom quality. Chien et. al (2010) also measured engagement in preschool classrooms across contexts of the day and found that levels of engagement varied in different classroom contexts and that children who had more quality instructional time with a teacher were better prepared for school. They also suggest that measuring engagement may be a better indicator of children's school readiness than classroom quality as measured by the ERS. Other studies of engagement in preschool found that children in classrooms who spent approximately equal time in teacher-directed group

activities and free play had higher engagement and better academic outcomes than children who spent most of the day in free play (Fuligni et al., 2012) and that children were more engaged when teachers addressed them individually than as part of a group (McWilliam, Scarborough, & Kim, 2003). Only a few studies of child engagement have included toddlers in their sample (McWilliam et al., 2003; Raspa et al., 2001; Ridley et al., 2000), although these studies did not include a rationale for including toddlers. More research needs to be done to better understand how engagement applies in toddler classrooms and how classroom context affects toddlers' engagement.

Measuring Engagement in Child Care Classrooms

One measure of engagement is the Engagement Check II (McWilliam, 1999). This is an environmental scan that considers all the children in the classroom at once, but it is still able to capture important variables of process quality obscured by more global measures by measuring children's engagement levels. When using the Engagement Check II, an observer observes the classroom for 15 seconds and then records the percentage of children engaged in activities or interactions. Studies using this measure have found that classrooms with lower global quality as measured by the ERS had lower levels of engagement (de Kruif, McWilliam, Ridley, & Wakely, 2000; Raspa et al., 2001; Ridley et al., 2000). These studies have only considered the composite ERS score and have not looked at how engagement relates to the subscales of the ERS.

Group engagement measures like the Engagement Check II are both effective and efficient and are able to provide immediate and relevant feedback related to quality, although they do not distinguish between levels of engagement in the same way that focal child measures can (Raspa et al., 2001; Ridley et al., 2000). Using

individual-child engagement observations is also an effective way to examine quality from the bottom-up perspective. For example, the Engagement Quality System III (E-Qual; McWilliam, 1998), a focal child measure, distinguishes between sophisticated, differentiated, focused, and unsophisticated engagement to determine if higher levels of engagement lead to better child outcomes (McWilliam et al., 1985). However, this type of methodology is less practical for applied purposes because of the time and expense associated with completing multiple observations.

Group measures can provide a picture of process quality by focusing on children's behavior and experience in the classroom and can help illuminate the relationship between quality and child outcomes. Therefore, group engagement measures can be used as a much-needed supplement to measures that focus on global quality, such as the ITERS-R, or on teacher behavior (Raspa et al., 2001; Ridley et al., 2000). Although they do not fully capture each child's experience because scores are based on the percentage of children engaged, it would not be possible for a classroom to score well if only a small number of children were engaged in activities or interactions, whereas this scenario may be possible with a more global measure of quality.

Chapter 3

METHODS

This study examines the relationship between global quality scores and levels of child engagement in toddler classrooms. Specifically, associations between global measures of quality using the ITERS-R and child engagement using the Engagement Check II are examined.

The study seeks to answer three primary research questions:

- What is the relationship between the percentage of children engaged and a classroom's ITERS-R score, and does that relationship change as quality increases?
- Do programs at different star levels have significant differences in engagement and ITERS-R scores?
- 3. Do engagement levels vary across different classroom contexts?

It was hypothesized that classrooms with higher classroom quality as measured by the ITERS-R have a higher mean percentage of child engagement and that there are differences in engagement and global quality by star level. It was also predicted that engagement levels vary across different classroom contexts, with higher levels of engagement expected during free play and lower levels during group times, transitions, and routines.

Context of the Study

This study took place within the context of Delaware Stars *for Early Success*, which is the statewide voluntary QRIS. Programs in Delaware Stars earn points through selecting standards and by achieving a threshold score on the ERS in selected classrooms. There are five star levels, and programs at higher star levels are eligible for a variety of financial incentives. Programs have the opportunity to request one practice ERS observation per star level to help them prepare for verification at a higher star level. Practice observations are full observations completed by reliable Delaware Stars assessors.

Because all of the participating programs were enrolled in Delaware Stars and the study utilized Delaware Stars assessors to complete ITERS-R classroom observations, certain Delaware Stars procedures impacted the methodology of this study. Some classroom assessments are completed by two Delaware Stars assessors to ensure inter-rater reliability. This study required two observers to be present in the classroom simultaneously to collect data on two different measures. Because it is a Delaware Stars policy to limit the number of observers in a classroom at a time to two, classrooms used for ERS inter-rater reliability were not included in this study. This policy also affected inter-rater reliability procedures for the engagement measure used. To ensure no more than two adults were collecting data in a classroom, observers used the on-campus child care center to conduct inter-rater reliability checks.

Sample

This study included 30 toddler classrooms in licensed child care centers participating in Delaware Stars. For this study, toddler classrooms were defined as classrooms where all children were at least 12 months and the majority of children were less than 36 months. Programs with toddler classrooms that requested an official verification or practice ITERS-R assessment from January to April of 2013 were eligible to participate in the current study. All programs with toddler classrooms requesting an assessment were asked to participate in the study with the exception of programs that were identified for inter-rater reliability purposes for Delaware Stars.

The primary researcher contacted each program that met the eligibility criteria and invited them to participate in the study. Of the 32 centers eligible to participate, 30 participated, 1 center declined to participate, and 1 center could not be reached prior to the Delaware Stars observation, which resulted in a participation rate of 94%. Participating classrooms received a \$25 gift card to a local classroom supply store as incentive for participation. One toddler classroom was observed from each participating program. All participating programs were a Star 2, Star 3, or Star 4 in the QRIS; no participating programs were at the highest star level.

52 adults participated in the study, which includes the lead teachers and assistant teachers who were present in the participating classrooms during the time of the observation. Table 1 shows participating teachers' demographic data, and Table 2 shows demographic data for the participating programs and classrooms.

Table 1

Teacher Demographic Data

Variable	N	%
Gender		
Male	0	0
Female	52	100
Race		
White	28	54
African American	17	33
Other	3	6
Ethnicity		
Hispanic	5	10
Not Hispanic	30	58
No answer	17	32
Education		
High School/GED	14	27
Some college credits	20	38
Associate's degree	7	13
Bachelor's degree	5	10
Some graduate credits	2	4
Graduate degree	4	8
Specialized Training		
CDA	3	6
ECE Associate's degree	4	8
ECE college credits	15	29
ECE bachelor's degree	2	4
ECE graduate credits	3	6
None of these	25	48
Current Position		
Curriculum Coordinator	2	4
Teacher	31	60
Assistant Teacher	16	31
Intern	3	6
Years of Experience		
0-5	26	50
6-10	14	27
11 or more	12	23

Table 2

Program and Classroom Demographic Data

Variable	Ν	%	
Program Type			
Full-Day	29	97%	
Part-Day	1	3%	
County			
New Castle	22	73%	
Kent	3	10%	
Sussex	5	17%	
Assessment Type			
Verification	17	57%	
Practice	13	43%	
Star Level			
Star 2	10	33%	
Star 3	11	37%	
Star 4	9	30%	
Number of Staff Present			
1	8	26%	
2	20	67%	
3 or more	2	7%	
Number of Children Present			
2-6	10	33%	
7-11	15	50%	
12 or more	5	17%	

Measures

The ITERS-R was used to measure global classroom quality, which is the observational tool currently used in Delaware's QRIS to assess infant and toddler classrooms where the majority of children are up to 36 months of age. The ITERS-R assessments were completed by Delaware Stars assessors who have earned at least 85% initial reliability on the tool, with inter-rater reliability checks conducted every

quarter. Delaware's QRIS does not include the Parents and Staff subscale of the ITERS-R when calculating the composite score. The omission of this subscale has become standard in large research studies and in other states' QRIS (Bisceglia et al., 2009; Malone, Kirby, Caronongan, Tout, & Boller, 2011), and this study excluded that subscale as well. Although some challenge the reliability and validity of the measure, the authors suggest that the ITERS-R has high reliability and validity, with a Cohen's Kappa of .55 for the scale without Parent and Staff, and high internal consistency (Environment Rating Scales Institute, n.d.; Harms et al., 2003).

The Engagement Check II (McWilliam, 1999) was used to measure group engagement and determine the overall percentage of children engaged. In this measure engagement is defined as "attention to or active participation in classroom activities as reflected by visual fixation, manipulation, vocalization, approach, or affect" (Raspa et al., 2001, pg. 214-215). Only behavior that was both developmentally and contextually appropriate was considered engagement. Using this measure, the observer counted the number of children engaged and nonengaged visible in one pass and recorded these numbers on a coding sheet. One observation was made every 15 seconds. The session score consisted of 60 observations made during a 15-minute session and yielded an average percentage of children engaged and nonengaged during the observation session. Six observation sessions were completed in each classroom, and the scores for each session were averaged to obtain the overall group engagement score.

The engagement measure was modified to capture information about six different classroom contexts—free play, whole group, small group, transition, mealtime, and personal care routines—as well as adult engagement with children to provide more in-depth information about how toddlers' engagement differs across

classroom context and when adults are engaged with them. For each observation interval, observers also recorded whether the interval took place inside or outside of the primary classroom. Locations outside the classroom that were observed included a playground, an indoor gross motor room, a hallway, and a bathroom that the whole class visited together that was not attached to the classroom. Each context was defined in a codebook after observers reached consensus on a definition. Definitions are listed in Table 3. The observers also adapted the coding sheet for the Engagement Check II to include contexts and adult engagement, as well as a cover sheet to record the number of children and adults present and brief notes about what occurred in the classroom during the observation cycle. The coding sheet can be found in the Appendix.
Table 3

Definitions and Common Examples of Contexts

Context	Definition and common examples
Free Play	Children have choice of materials, location in classroom, and
	playmate
	Examples:
	Classroom center time
	Play in self-selected groups
Whole Group	Children participate in a teacher-initiated activity intended for whole
	group participation
	Examples:
	Circle time with a story and music
	Whole group art project
Small Group	Children participate in a teacher-initiated activity intended for a small
	group of children
	Examples:
	Four condition completing an art activity
Maaltima	Children esting a meal or speak boring when food is served
Meannie	Examples:
	Lunch time
	Morning snack
Personal Care	Children are involved in activities and routines associated with
Routines	personal care
itoutilos	Examples:
	Teacher changing a child's diaper
	Child washing hands after using the bathroom
Transition	Children are involved in a teacher-initiated transition between two
	activities and/or two locations
	Examples:
	Cleaning up from free play to being group time
	Putting on coats and lining up to go outside
Adult	Adults are engaged with at least one child in the same context in
Engagement	which the child or children are present
	Examples:
	Building blocks with two children during free play
	Reading a story to the class during whole group time

Inter-rater reliability for the engagement measure was calculated by taking the lowest group engagement score for a session between two observers and dividing it by the highest engagement score, which is the procedure used in other published studies using this measure (Raspa et al., 2000; Ridley et al., 2000). Observers who conducted the measure were trained until they reached inter-rater reliability of at least 90% for overall engagement and at least 90% for each classroom context observed. They completed six inter-rater reliability checks at the on-campus child care center spread out at approximately equal intervals throughout the four month data collection period, with inter-rater reliability of 95% for overall engagement and at least 90% for each classroom context.

Procedures

The modified Engagement Check II and the ITERS-R were conducted in classrooms simultaneously by two different observers. Lead and assistant teachers in participating classrooms signed informed consent forms and completed demographic forms. The observer who completed the engagement measure also recorded the number and gender of children and adults present and took brief qualitative notes at the start and end of each observation cycle describing what was happening in the classroom.

Six 15-minute sessions of the Engagement Check II were completed during one visit, which resulted in 360 engagement data points, as well as a mean engagement percentage for each observation session and an overall percentage of engagement for the classroom. Observers wore a headphone connected to an audio player playing a sound file that altered them to the end of each 15-second interval and each 15-minute observation cycle. Observers also captured information about classroom context. They recorded how many children were present in and engaged in the six different classroom contexts during each 15-second interval using the coding sheet found in the Appendix. Additionally, they recorded how many adults were engaged with children in each context. It was possible for more than one classroom context to be occurring during a 15-second interval. For example, some children in the classroom may have been participating in free play while others were participating in a small group activity. In these instances, the observer captured each classroom context occurring and the number of children present and engaged in each context.

The observers completing the engagement measure used a copy of the classroom's schedule to strategically select six 15-minute observation cycles that occurred while the assessor conducted the ITERS-R assessment. The ITERS-R assessment began at approximately 8:30 am and ended when the classroom began nap time, which varied from approximately 11:00 am to 1:30 pm. Observers selected engagement observation cycle times to try to see as many classroom contexts as possible throughout the observation period and to allow for rest time between each observation cycle. All parts of the schedule were eligible to be observed, including outdoor time and times when the lead teacher may have been out of the room in order to best capture children's experiences.

Observers were able to observe free play, mealtime, and transitions in all 30 classrooms. Whole group and small group did not occur in all 30 classrooms; these contexts were observed in 27 and 10 classrooms respectively. Personal care was not able to be observed in 4 classrooms because diapering and hand washing took place in a separate room with a closed door.

The mean, standard deviation, minimum, and maximum number of observations that were observed for each context for the 30 classrooms are listed in Table 4. Although 360 timed observations of engagement occurred in each classroom, the total number of observations when combining the six classroom contexts was higher than 360 in each classroom due to recording co-occurring contexts separately. Free play was the most frequently observed context, and small group was least frequently observed. Of the 10800 observation points of engagement that occurred across the 30 classrooms, 1 context was observed for 8558 (79.2%) of the observations, 2 contexts were observed for 2097 (19.4%) observations, and 3 or 4 contexts were observed during 145 (1.4%) observations. Adult engagement was captured for each of the 10800 observations. 9183 of the 10800 observations occurred inside the primary classrooms, and 1617 took place outside the classroom, including hallways, playgrounds, bathrooms not attached to the classroom, and gross motor rooms.

Table 4

Frequency of Observations by Context

Classroom				
context	Mean	SD	Minimum	Maximum
Overall	439.87	53.94	361	560
engagement				
Free play	185.43	62.78	23	306
Whole group	51.73	44.06	0	198
Small group	13.47	24.88	0	98
Mealtime	79.30	29.09	27	122
Personal care	36.43	23.64	0	76
Transition	73.50	49.17	8	229

Chapter 4

RESULTS

Table 5 presents means, standard deviations, minimums, and maximums for the 30 classrooms observed.

Table 5

Descriptive Statistics of Observational Data by Classroom

Variable	Mean	Std. Dev	Minimum	Maximum
Children Present	7.80	3.98	2	20
Engagement	58.39	11.47	34.21	74.77
ITERS Composite Score	3.76	.98	1.67	5.03
Space/Furnishings	3.79	1.01	1.80	5.60
Personal Care	1.87	.67	1.00	3.33
Listening and Talking	4.97	1.89	1.00	7.00
Activities	3.91	1.05	1.60	5.56
Interactions	4.83	1.73	1.25	7.00
Program Structure	4.43	1.59	1.33	7.00

The Relationship between Engagement and Global Quality

Using a Pearson correlation, it was found that the relationship between percentage of children engaged and a classroom's ITERS-R score was statistically significant and reflects a large effect size (r = .607, p = .001). Table 6 shows the correlations between engagement and the ITERS-R composite score and subscale scores. There was statistically significant positive correlation between each subscale of the ITERS-R and engagement with the exception of Personal Care Routines, although the correlation was the strongest with the composite score. The subscales with the strongest correlations with engagement were Program Structure, Activities, and Interactions. The relationship between adult engagement and the ITERS-R composite score was not significant (r = .351, p = .057).

Table 6

Pearson Correlations between ITERS-R Subscales and Overall Engagement

ITERS-R	Engagement
ITERS-R Composite Score	.607**
Space and Furnishings	.443*
Personal Care Routines	.227
Listening and Talking	.426*
Activities	.595**
Interactions	.556**
Program Structure	.596**
* <i>p</i> <.05; ** <i>p</i> <.01	

The scatter plot in Figure 1 shows the positive relationship between ITERS-R composite score and the percentage of child engagement.



Figure 1 Scatter Plot of Relationship between ITERS-R and Child Engagement

Differences in Engagement between Classrooms by Global Quality

To determine if there was a difference in engagement between classrooms based on global quality, the sample was divided into two groups, low quality and midto high quality, based on each classroom's composite score on the ITERS-R. Classroom that scored a 4.0 or higher were considered mid- to high quality, and classrooms scoring below a 4.0 were considered low quality. This cut-off point was determined considering the median score of the sample, which was 4.09, and the mean, which was 3.76, as well as the design for the scale. In the scoring scale of the ITERS-R, a 4.0 falls between "minimal" quality and "good" quality (Harms et al., 2006). The dependent variable of engagement was on the interval scale of measurement. Therefore, data were analyzed using an independent samples *t*-test. Table 7 presents means and standard deviations for the dependent variable of engagement separately for the two groups. Results showed a statistically significant difference between classrooms with low quality classrooms obtaining lower engagement scores and those that were mid- to high-quality (t = -2.50, df [28], p = .02). The effect size is large (Cohen, 1998; d = -0.92) and represents a substantive difference between the two groups. These results suggest that there is a significant different in engagement between classrooms scoring low and those scoring mid-to-high on the ITERS-R.

Table 7

Differences in Engagement by Mid-to-High and Low ITERS-R Scores

Group	Μ	SD
Low Quality	52.89	12.02
(N=13)		
Mid-to-High Quality	62.61	9.32
(N=17)		

Differences in Engagement and Global Quality by Star Level

The second research question considers whether a program's rating in the QRIS can be differentiated by a classroom's engagement and ITERS-R scores. To answer this question data was analyzed using a one-way between-subjects multivariate

analysis of variance (MANOVA). Dependent variables were the mean percentage of engagement and the ITERS-R composite score, and the independent variable was the program's star level. Distributional statistics for the scores on the two measures are presented separately for each group in Table 8.

Table 8

	Sta	Star 2		Star 3		Star 4	
Scores	(N=	(N=10)		:11)	(N=9)		
	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	<u>M</u>	<u>SD</u>	
Engagement	49.64	12.28	60.32	8.84	65.76	6.91	
ITERS-R	2.95	1.11	3.92	0.70	4.47	0.26	

Mean Engagement and ITERS-R Scores by Star Level

A priori power was assessed for the MANOVA. Pillai's trace was utilized as the multivariate test statistic because it is robust, especially when working with small sample sizes (Bray & Maxwell, 1985; Olson, 1976; Stevens, 2002). The significance level was set to p = .05 for the power analysis, as per standard scientific conventions. Power was set to .80, meaning there would be an 80% probability of reaching statistical significance if the obtained sample differences were present in the population. The power analysis was completed twice using both a medium effect size $(f^2 = .15)$ and a large effect size $(f^2 = .30; Cohen, 1988)$. Results from the power analysis showed 45 cases would be necessary for the multivariate test with a medium effect and 24 cases with a large effect. Although Box's test was not significant, which indicates that the variancecovariance matrices of the dependent variables across the three star levels are approximately equal, Pillai's trace was employed to assess the multivariate effect due to the small sample sizes (Stevens, 2002; Tabachnick & Fidell, 2007). The multivariate effect showed that engagement and ITERS-R scores differed significantly based on a program's star level (Pillai's Trace = .467, F = 4.114, df [4,54], p = .006) and represented a medium-to-large effect size (partial $\eta^2 = .234$; Cohen, 1988).

Following this, univariate ANOVAs were conducted to determine how engagement and ITERS-R composite scores each differentiate star level. Multiple comparison tests were completed to help isolate where the differences were located. The ANOVA showed that engagement (F = 6.938, p = .004) and ITERS-R composite score (F = 9.234, p = .001) both significantly contributed to the differences in star level. Games-Howell was selected as the multiple comparison because the ITERS-R composite scores violated the assumption of homogeneity of variances (Levene Statistic F=9.685, p = .001). Results of the multiple comparison tests showed that the difference between Star 2 and Star 4 programs was statistically significant for both engagement (p = .008) and ITERS-R composite score (p = .005). The differences between Star 2 and Star 3 and Star 4 were not statistically significant for either engagement or ITERS-R composite score. This is likely because of the small sample sizes in each group.

Engagement by Classroom Context

Two methods were employed to examine engagement data by context. Descriptive statistics for engagement by classroom are presented in Table 9, and descriptive statistics for engagement across all observation points are presented in Table 10. In both ways of examining the data, transition was the context with the lowest engagement, and mealtime had the highest engagement.

Table 9

Descriptive	Statistics of	of	Engagement	Contexts	by	Classroom
· · · · · · · · · · · · · · · · · · ·		- J			- 2	

	Frequency		Mean	
Context	Ν	N % En		SD
Total Engagement	30	100.00	58.39%	11.47
Free Play	30	100.00	63.76%	10.13
Whole Group	27	90.00	50.38%	18.31
Small Group	10	33.33	63.61%	20.72
Mealtime	30	100.00	77.37%	9.21
Personal Care	26	86.67	55.87%	22.86
Transition	30	100.00	34.89%	12.66
Adult Engagement	30	100.00	67.07%	13.79

Table 10

Descriptive Statistics of Engagement Contexts across All Observation Intervals

	Frequency		Mean	
Context	Ν	%	Engagement	SD
Total Engagement	10800	100.00	58.74%	27.79
Free Play	5563	51.51	65.52%	23.63
Whole Group	1552	14.37	48.56%	30.32
Small Group	404	3.74	60.79%	34.82
Mealtime	2379	22.03	76.88%	25.97
Personal Care	1093	10.12	56.77%	47.35
Transition	2205	20.42	31.71%	30.18
Adult Engagement	10800	100.00	69.02%	34.01

Correlations were completed between the ITERS-R composite score and engagement in each classroom context to determine if some contexts had a stronger relationship with the ITERS-R score than others. Results are shown in Table 11. Three contexts had a significant positive relationship with the ITERS-R composite score: whole group, mealtime, and transition.

Table 11

	ITERS-R Composite
Context	Score
Engagement	.607**
Adult Engagement	.328
Free Play Engagement	.125
Whole Group Engagement	.506**
Small Group Engagement	.297
Mealtime Engagement	.500**
Personal Care Engagement	.344
Transition Engagement	.502**
*n < 05 **n < 01	

Correlations between Engagement by Context and ITERS-R Composite Score

p* < .05, *p* < .01

Although adult engagement was not significantly related to the ITERS-R composite score, it was significantly related to children's overall engagement (r = .045, p = .001). This suggests that during intervals when adults were engaged with children, child engagement levels were higher.

Because each observation was coded as occurring either inside or outside the primary classroom, it was also possible to look at the differences between engagement in and out of the classroom using an independent samples *t*-test. The mean engagement for observations in the classroom was 59.44% (SD=27.77), and mean engagement out of the classroom was 54.75% (SD=27.79). Results showed a statistically significant difference between engagement in the classroom and out of the classroom (t = 6.27, df [10798], p = .001), although the effect size is relatively small (Cohen, 1988; d = 0.17). These results suggest that there is a significant difference in engagement inside and out of the primary classroom and that there are higher levels of group engagement in the classroom. An independent samples *t*-test was also completed to determine if there were significant differences between observation intervals where one context occurred and where two or more contexts were cooccurring. The differences between the groups was not statistically significant (t = -.393, df [3735.30], p = .762).

Chapter 5

DISCUSSION

The present study confirms that engagement and global quality as measured by the ITERS-R are positively correlated and expands that finding with a more in-depth picture of how the two constructs are related. The addition of classroom contexts and adult engagement with children into the observations of engagement allowed for exploration into how engagement varies by context and with adult support in child care classrooms and how these factors relate to global quality. Further, by including programs participating in Delaware's QRIS, it was possible to examine the way engagement and global quality vary by star level. A focus on toddler classrooms was another unique component of this study and helps to answer the important question of how to improve the quality of child care for very young children (Ackerman & Barnett, 2009; NICHD, 2002). The findings also support the value of viewing engagement as a proximal process to improve children's everyday experiences (Bronfenbrenner & Morris, 1998; Bronfenbrenner & Evans, 2000).

Engagement and Global Quality

Similar to previous studies of group engagement and classroom quality (Raspa et al., 2001; Ridley et al., 2000), the data indicate that there is a positive relationship between a classroom's score on the ITERS-R and the mean percentage of group engagement. Children in classrooms that had lower global quality were more likely to spend time non-engaged or acting in ways that were not contextually appropriate,

while children in classrooms with higher global quality spent more time actively engaged with adults, peers, and materials, which has been linked to positive child outcomes (Fuligini et al., 2012).

These findings suggest that engagement can be used as one indicator of classroom quality. However, it is clear from the large standard deviations in engagement data that there is wide variation in engagement across classrooms. This supports previous research that global measures of quality may not be sufficient in capturing the quality of children's experiences in child care (Bisceglia et al., 2009; Thomason & La Paro, 2009; Wiltz & Klein, 2001) and supports claims that additional research is needed that utilizes a bottom-up approach to measuring quality (Ceglowski & Bacigalupa, 2002; Hallam et al., 2009; Katz, 1993). Measuring engagement differs from measuring global quality in that it directly incorporates the individual experiences of children during each observation point. Findings from this study indicate that measuring the engagement experiences of each child in a classroom in addition to using a global measure of quality provides different information than would be obtained from using only a global measure.

This study examined the relationship of the different subscales of the ITERS-R with a classroom's mean group engagement, although it is recognized that the subscales may not each measure distinct areas of quality (Bisceglia et al., 2009; Hestenes et al., 2007). Personal Care Routines was the only subscale not significantly correlated to group engagement. This is likely due in part to the low scores on the Personal Care Routines subscale across the sample (ranging from 1.00 to 3.33 with a mean of 1.67). It also suggests that higher scores on this subscale may not foster higher levels of child engagement.

Differences by Star Level

There were significant differences in programs at Star 2, Star 3, and Star 4 as a function of their group engagement and their ITERS-R composite score, with increasing mean ITERS-R scores and percentages of group engagement at each star level. The results of the one-way ANOVAs that followed the MANOVA revealed that differences were significant for both engagement and ITERS-R composite score.

It is expected that differences would be significant for the ITERS-R, as it is currently the classroom observation measure used in toddler classrooms for rating in the QRIS. Although only one-third of the classrooms in each age group are selected to be assessed, the expectation is that each classroom would be able to score at least the minimum score threshold required for each star level. There appear to be slight differences in engagement and in ITERS-R composite scores between each star level, although due to the small number of classrooms at each star level, the multiple comparisons were not powerful enough to detect differences that may have been present between Star 2 and Star 3 and between Star 3 and Star 4, although differences were statistically significant between Star 2 and Star 4. These findings also point to the possibility of including a measure of engagement in the QRIS to measure and rate classroom quality, as well as the larger value of including multiple measures to better capture children's experiences (Cassidy et al., 2005; Denny et al., 2012; Dickinson, 2002).

Engagement and Classroom Context

Through capturing the contexts in which children spent their time, it is possible to gain a more dynamic picture of how children's engagement varies based on the classroom context in which they are participating. Descriptive analysis of the data confirms that engagement does vary by context (Chien et al., 2010; Downer et al., 2007; Tonyan & Howes, 2003). Although mealtime had the highest mean percentage of engagement, this does not necessarily indicate that children were frequently engaged in sophisticated ways during mealtimes, as the environmental scan of engagement does not distinguish between levels of engagement in the same way focal child measures do (McWilliam, 1998). Engagement during whole group was higher than hypothesized, which is likely also because of the way engagement was measured. A measure that distinguished between levels of engagement may reveal lower levels of sophisticated engagement in mealtime and whole group activities.

Just as there is wide variation in overall engagement by classroom, there is also wide variation in engagement by context for classrooms. Future research that includes a more in-depth examination of the contexts may help reveal reasons why this variation is so great, although findings suggest that young children are more engaged in contexts where they are more likely to have hands-on, play-based, and child-guided experiences, which is in agreement with the National Association of the Education of Young Children's position statement on Developmentally Appropriate Practice (Bredekamp & Copple, 1997).

The three contexts that showed a significant positive relationship to ITERS-R composite scores were mealtime, whole group, and transition, which suggests that classrooms that have higher engagement in these three contexts may also have higher global quality. This has implications for improving teacher practice and for quality improvement efforts. It is possible that decreasing wait times during meals and transitions, times when children are generally not engaged, is one way to improve global quality and to make toddler classrooms more developmentally appropriate

through providing more hands-on experiences (Bredekamp & Copple, 1997). Also, ensuring whole group experiences for young children are engaging and developmentally appropriate may be another strategy for improving global quality.

The especially low engagement during transitions shows that there is much room for improving young children's experiences with transitions in child care. The lower engagement outside the classroom also points to room to improve the experiences toddlers have outside of their classroom, which in this study was primarily outdoor play. It is notable that free play was not significantly correlated with ITERS-R composite score. This may support the finding that the Environment Rating Scales place more emphasis on structural quality compared to process quality (Cassidy et al., 2005).

Limitations

The primary limitations in this study relate to the sample. The sample size of 30 classrooms is small and limits the power for completing statistical analyses at the classroom level. Also, the study used a convenience sample of programs that participated in the voluntary QRIS and had requested a classroom assessment. For this reason, findings may not be generalizable to toddler classrooms in all child care centers. Another limitation of the study was a relatively small range of ITERS-R scores. The mean score was 3.76, which falls between "minimal" and "good" quality according to the language used in the scale (Harms et al., 2006). No classroom scored above a 5.03 on the ITERS-R, so it is not possible to draw conclusions about the relationship between engagement and global quality when global quality is high. Low composite scores on the ITERS-R, while a limitation in this study, are fairly consistent with national and state data from QRIS, with ITERS-R scores tending to be lower than

scores on the ECERS-R and often falling below the "good" level as defined in the scale (Malone et al., 2011). The mean ITERS-R score for this study is also similar to the state average ITERS-R score for verifications in the QRIS, which is a 3.91. Because there were no Star 5 programs included in the sample, it is not possible to conclude if the pattern of differences in engagement and ITERS-R scores applies at this highest star level as well. Finally, although a strength of the group engagement scan is the ability to capture information about every child in the classroom at each observation point, the measure does not distinguish between higher and lower levels of engagement. Including a focal child measure of engagement could provide deeper information about how engagement relates to global quality and varies by context and by individual child (McWilliam et al., 1985).

Implications and Future Directions

The findings from this study provide rationale for exploring the inclusion of engagement measures in quality rating in addition to global measures of quality to better capture children's experiences in the classroom. The very low levels of engagement during transition suggest that teachers could use support in decreasing the time toddlers spend in transition and improving children's engagement during these transitions. There is also evidence that focusing classroom improvement efforts on improving engagement during whole group times, mealtimes, and transitions may positively impact ITERS-R scores and help programs move to higher levels in the QRIS. Focusing on engagement and strategies for increasing it in group and individual professional development efforts, such as coaching and technical assistance, is one strategy for equipping teachers to improve children's engagement in the classroom.

Future research in this area should include a wider sample of programs, including those not participating in the QRIS and those at the highest star level. Having a larger sample of classrooms would also help detect significant differences between star levels. Including a measure of individual children's engagement in future research would also be helpful to see if individual children's engagement levels are similar to the mean engagement of all children in the classroom or if there is significant variation between individual children. A more individual measure of engagement would also allow for richer information about the type of engagement that occurs.

These findings suggest that a closer consideration of the role engagement plays in classroom quality is worthwhile and that it is important to look beyond using only one measure of global quality to in order to capture the quality of young children's everyday experiences in child care more completely (Denny et al., 2012; Dickinson, 2002). It also reveals that it is important to consider in which classroom contexts children are spending their time and what options are available to them in those contexts, as well as the ways in which a focus on improving the quality of engagement in one context may have implications for overall classroom quality.

REFERENCES

- Ackerman, D. J., & Barnett, W. S. (2009). NIEER Policy brief: Does preschool education policy impact infant/toddler care? Retrieved from http://nieer.org/resources/policybriefs/21.pdf
- Belsky, J. (1988). The "effects" of infant day care reconsidered. *Early Childhood Research Quarterly, 3,* 235-272.
- Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., Owen, M. T., & The NICHD Early Child Care Research Network. (2007). Are there long-term effects of early child care? *Child Development*, 78, 681-701. doi:10.1111/j.1467-8624.2007.01021.x
- Bisceglia, R., Perlman, M., Schaack, D., & Jenkins, J. (2009). Examining the psychometric properties of the Infant-Toddler Environment Rating Scale-Revised Edition in a high-stake context. *Early Childhood Research Quarterly*, 24, 121-132.
- Booren, L. M., Downer, J. T. & Vitiello, V. E. (2012). Observations of children's interactions with teachers, peers, and tasks across preschool classroom activity settings. *Early Education and Development*, 23(4), 517-538.

- Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance*. Newbury Park, CA: Sage.
- Bredekamp, S., & Copple, C. (Eds.). (1997). Developmentally appropriate practice for early childhood programs. Washington, DC: National Association for the Education for Young Children.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.
- Bronfenbrenner, U. (1992). Ecological systems theory. In R. Vasta (Ed.), Six theories of child development: Revised formulations and current issues (pp. 187-249).
 London, UK: Jessica Kingsley.
- Bronfenbrenner, U. (1999). Environments in developmental perspective: Theoretical and operational models. In S. Friedman & T. Wachs (Eds.), *Measuring environment across the life span: Emerging methods and concepts* (pp. 3-28).
 Washington, DC: American Psychological Association Press.
- Bronfenbrenner, U. & Evans, G. W. (2000). Developmental science in the 21st century: Emerging questions, theoretical models, research designs and empirical findings. *Social Development*, *9*(1), 115-125.

Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental process. In
W. Damon (Series Ed.) & R. M. Lerner (Vol. Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development* (5th ed., pp.
993–1028). New York, NY: Wiley.

Burchinal, M. R., Nabors, L., Bryan, D., & Roberts, J. (1996). Quality of center infant care and infant cognitive, language, and social development. *Child Development*, 67, 606-620.

- 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(3), 345-360.
- Caughy, M. O., DiPietro, J. A., Strobino, D. M. (1994). Day-care participation as a protective factor in the cognitive development of low-income children. *Child Development*, 65, 457-471. doi:10.1111/j.1467-8624.1994.tb00763.x
- Ceglowski, D. A. (2004). How stake holder groups define quality in child care. *Early Childhood Education Journal*, *32*(2), 101-111.
- Ceglowski, D., & Bacigalupa, C. (2002). Four perspectives on child care quality. *Early Childhood Education Journal, 30*(2), 87-92.

- Chien, N. C., Howes, C., Burchinal, M., Pianta, R. C., Ritchie, S., Bryant, D. M., ... Barbarin, O. A. (2010). Children's classroom engagement and school readiness gains in prekindergarten. *Child Development*, 81(5), 1534-1549.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Cryer, D. (1999). Defining and assessing early childhood program quality. *Annals of the American Academy of Political and Social Science*, *563*, 39-55.
- Davies, D. (2004). Risk and protective factors: The child, family, and community contexts. *Child development: A practioner's guide* (2nd ed., pp. 61-108). New York: The Guilford Press.
- de Kruif, R. E. L., McWilliam, R. A., Ridley, S. M., & Wakely, M. B. (2000).
 Classification of teachers' interaction behaviors in early childhood classrooms.
 Early Childhood Research Quarterly, 15(2), 247-268.
- Denny, J. H., Hallam, R., & Homer, K. (2012). A multi-instrument examination of preschool classroom quality and the relationship between program, classroom, and teacher characteristics. *Early Education & Development*, 23(5), 678-696.
- Dickinson, D. K. (2002). Shifting images of developmentally appropriate practice as seen through different lenses. *Educational Research*, *31*(1), 26-32.

- Downer, J. T., Rimm-Kaufman, S. E., & Pianta, R. C. (2007). How do classroom conditions and children's risk for school problems contribute to children's behavioral engagement in learning? *School Psychology Review*, 36(3), 413-432.
- Dunn, L. (1993). Proximal and distal features of day care quality and children's development. *Early Childhood Research Quarterly*, *8*, 167-192.
- Egeland, B., & Hiester, M. (1995). The long-term consequences of infant day care and mother–infant attachment. *Child Development*, *66*, 474–485.
- ERSi. (n.d.) *Reliability and validity of the infant/toddler environment rating scale*. Retrieved from http://www.ersi.info/PDF/ReliabilityIters.pdf
- Flanagan, K., & West, J. (2004). Children born in 2001: First results from the base year of the Early Childhood Longitudinal Study, Birth Cohort (ECLS–B) (NCES 2005–036). Washington, DC: National Center for Education Statistics. Retrieved from http://www.childcareresearch.org/location/ccrca4623
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59-109.

- Greenwood, C. R., Horton, B. T., & Utley, C. A. (2002). Academic engagement: Current perspectives on research and practice. *School Psychology Review*, *31*(3), 328-349.
- Fuligni, A. S., Howes, C., Huang, Y., Hong, S. S., & Lara-Cinisomo, S. (2012). Activity settings and daily routines in preschool classrooms: Diverse experiences in early learning settings for low-income children. *Early Childhood Research Quarterly*, 27, 198-209.
- Girolametto, L. & Weitzman, E. (2002). Responsiveness of child care providers in interactions with toddlers and preschoolers. *Language, Speech, and Hearing Services in Schools, 33*(4), 268-281.
- Hallam, R., Fouts, H., Bargeen, K., & Caudle, L. (2009). Quality from a toddler's perspective: A bottom-up examination of classroom experiences. *Early Childhood Research and Practice*, 11(2).
- Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first-grade classroom make a difference for children at risk of school failure? *Child Development*, 76, 949–967.
- Harms, T., Clifford, R., & Cryer, D. (1998) *Early childhood environment rating scale–revised*. New York: Teachers College Press.

- Harms, T., Cryer, D., & Clifford, R. (1990). *Infant/toddler environment rating scale*. New York: Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R. (2003). *Infant/toddler environment rating scale– revised*. New York: Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R. (2007). Family child care environment rating scale-revised. New York: Teachers College Press.
- Harms, T., Jacobs, E., White, D. (2006). School-age care environment rating scale.New York: 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. doi:10.1080/10409280701283106
- Helburn, S. W. (1995). Cost, quality, and child outcomes in child care centers: Technical report. Denver, CO: Dept of Economics, Center for Research in Economic and Social Policy, University of Colorado.
- Hestenes, L. L., Cassidy, D. J., Hegde, A. V., & Lower, J. K. (2007). Quality in inclusive and noninclusive infant and toddler classrooms. *Journal of Research in Childhood Education*, 22, 69-84. doi:10.1080/02568540709594613

- Howes, C., Burchinal, M., Pianta, R. C., Bryant, D., Early, D., Clifford, R., &
 Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in
 pre-kindergarten programs. *Early Childhood Research Quarterly*, 23(1), 27-50.
- Howes, C., Phillips, D. A., & Whitebook, M. (1992). Thresholds of quality:Implications for the social development of children in center-based child care. *Child Development*, 63, 449-460.
- Katz, L. G. (1993). Multiple perspectives on the quality of early childhood programs. ERIC Document Reproduction Service No: ED355 041.
- Katz, L. G. (1994). Perspectives on the quality of early childhood programs. *Phi Delta Kappan*, 76(3), 200-205.
- Knitzer, J., & Lefkowitz, J. (2006). Pathways to early school success—issue brief no. 1: Helping the most vulnerable infants, toddlers, and their families. Retrieved from http://www.nccp.org/publications/pdf/text_669.pdf
- Kreader, J. L., Ferguson, D., & Lawrence, S. (2005a). Infant and toddler child care arrangements. *Child Care & Early Education Research Connections: Research-to-Policy Connection No. 2.* Retrieved from http://www.nccp.org/publications/pdf/text_628.pdf

Kreader, J. L., Ferguson, D., & Lawrence, S. (2005b). Infant and toddler child care quality. *Child Care & Early Education Research Connections: Research-to-Policy Connection No. 2.* Retrieved from http://childcareresearch.org/childcare/resources/6872/pdf

La Paro, K. M., Thomason, A. C., Lower, J. K, Kinter-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 and Practice, 14*(1). Retrieved from http://ecrp.uiuc.edu/v14n1/laparo.html?goback=.gde_1807282_member_12831 4736

- 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). *Life in preschool: Final report* (Vol. 1). Cambridge, MA: Abt Associates.
- Malone, L., Kirby, G., Caronongan, P., Tout, K., & Boller, K. (2011). *Measuring quality across three child care quality rating and improvement systems: Findings from secondary analyses* (OPRE Report # 2011-30). Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

- 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(3), 732-749.
- McWilliam, R. A. (1998). *Engagement Quality Observation System III (E-Qual)*. Chapel Hill, NC: Frank Porter Graham Child Development Center.
- McWilliam, R. A. (1999). *Engagement Check II*. Chapel Hill, NC: University of North Carolina, FPG Child Development Institute.
- McWilliam, R. A., & Bailey, D. B. (1995). Effects of classroom social structure and disability on engagement. *Topics in Early Childhood Special Education*, 15, 123-147.
- McWilliam, R. A., Scarborough, A. A., & Kim, H. (2003). Adult interactions and child engagement. *Early Education and Development*, 14, 7-28. doi:10.1207/s15566935eed1401_2
- McWilliam, R. A., Trivette, C. M., & Dunst, C. J. (1985). Behavior engagement as a measure of the efficacy of early intervention. *Analysis and Intervention in Developmental Disabilities*, *5*, 59-71.

- Melhuish, E. C. (2001). The quest for quality in early day care and preschool experience continues. *International Journal of Behavioral Development*, 25, 1-6. doi:10.1080/01650250042000492
- Mitchell, Anne 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 Success by Six.
- National Association of Child Care Resource & Referral Agencies (NACCRRA). (2009). Comparison of quality rating and improvement systems (QRIS) with Department of Defense standards for quality. Retrieved from http://www.naccrra.org/sites/default/files/default_site_pages/2011/army_qris_s tate_report_081709.pdf
- NICHD Early Child Care Research Network. (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly, 11, 269-306.*
- NICHD Early Child Care Research Network (1997). The effects of infant child care on infant-mother attachment security: Results of the NICHD Study of Early Child Care. *Child Development*, 68, 860–879.
- NICHD Early Child Care Research Network. (2002). Child-care structure→process→outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science*, *13*(3), 199-206.

- National Scientific Council on the Developing Child (2004). *Children's emotional development is built into the architecture of their brains: Working paper no.* 2. Retrieved from http://www.developingchild.harvard.edu
- Odom, S. L., Vitztun, J., Wolery, R., Lieber, J., Sandall, S., Hanson, M.J. ... Horn, E. (2004). Preschool inclusion in the United States: A review of research from an ecological systems perspective. *Journal of Research in Special Education Needs*, *4*, 17-49. doi:10.1111/J/1471-3802.2004.00016.x
- Olson, C. L. (1976). On choosing a test statistic in MANOVA. *Psychological Bulletin*, 83, 579-586.
- Phillips, D., Mekos, D., Scarr, S., McCartney, K., & Abbott-Shim, M. (2000). Within and beyond the classroom door: Assessing quality in child care centers. *Early Childhood Research Quarterly*, 15(4), 475-496.
- Pianta, R. C., La Paro, K. M., & Hamre, B. (2008) *Classroom assessment scoring system (CLASS). Pre-K version.* Baltimore: Brookes.
- Raspa, M. J., McWilliam, R. A., & Ridley, S. M. (2001). Child care quality and children's engagement. *Early Education and Development*, *12*, 209-224. doi:10.1207/s15566935eed1202_3

- Ridley, S. M., McWilliam, R. A., & Oates, C. S. (2000). Observed engagement as an indicator of child care program quality. *Early Education and Development*, 11, 133-146. doi:10.1207/s15566935eed1102_1
- Scarr, S., Eisenberg, M., & Deater-Deckard, K. (1994). Measurement of quality in child care centers. *Early Childhood Research Quarterly*, *9*, 131-151.
- Shonkoff, J. P., & Phillips, D. A. (2000). From neurons to neighborhoods: The science of early childhood development. Washington, DC: National Academy Press.
- Shulman, K., Matthews, H., Blank, H., & Ewen, D. (2012). A count for quality: Child care center directors on rating and improvement systems. Retrieved from http://www.clasp.org/admin/site/publications/files/ACountforQualityQRISRep ort.pdf
- Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.): Mahwah, New Jersey: Lawrence Erlbaum.
- 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: Trentham Books.
- Tabachnick, B. G., & Fidell, L. S. (2007). Using multivariate statistics (5th. ed.).Boston: Pearson.

- Thomason, A. C., & La Paro, K. M. (2009). Measuring the quality of teacher-child interactions in toddler child care. *Early Education and Development*, 20, 285-304. doi:10.1080/10409280902773351
- Tonyan, H. A., & Howes, C. (2003). Exploring patterns in time children spend in a variety of child care activities: Associations with environment quality, ethnicity, and gender. *Early Childhood Research Quarterly*, 18, 121-142.
- Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G., & Boller, K. (2010). Compendium of quality rating systems and evaluations. Washington, DC: Child Trends & Mathematica Policy Research.
- Tout, K., Zaslow, M., Halle, T., & Forry, N. (2009). Issues for the next decade of quality rating and improvement systems. Retrieved from http://www.childtrends.org/Files//Child_Trends-2009_5_19_RB_ QualityRating.pdf
- Vandell, D. L., & Wolfe, B. (2000). Child care quality: Does it matter and does it need to be improved? Institute for Research on Child Poverty: University of Wisconsin-Madison.
- Wiltz, N. W., & Klein, E. L. (2001). "What do you do in child care?" Children's perceptions of high and low quality classrooms. *Early Childhood Research Quarterly*, 16, 209-236.

Appendix A

ENGAGEMENT CODING SHEET

	FP	WG	SG	GM	м	PC	Т
1							
2							
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
Appendix **B**

IRB LETTER

COLLABORATIVE INSTITUTIONAL TRAINING INITIATIVE (CITI) RESPONSIBLE CONDUCT OF RESEARCH CURRICULUM COMPLETION REPORT Printed on 11/25/2013

LEARNER DEPARTMENT EMAIL INSTITUTION EXPIRATION DATE Alison Hooper (ID: 3816934) Human Development and Family Studies 7174218426 seefeidt@udel.edu University of Delaware

SOCIAL AND BEHAVIORAL RESPONSIBLE CONDUCT OF RESEARCH COURSE 1. COURSE/STAGE: RCR/1 PASSED ON 10/13/2013 REFERENCE ID: 11500385

REQUIRED MODULES	DATE COMPLETED	SCORE
The CITI Course in the Responsible Conduct of Research	10/13/13	No Quiz
Introduction to the Responsible Conduct of Research	10/13/13	No Quiz
Introduction to Research Misconduct	10/13/13	No Quiz
Research Misconduct (RCR-SBE)	10/13/13	3/5 (60%)
Data Management (RCR-SBE)	10/13/13	5/5 (100%)
Authorship (RCR-SBE)	10/13/13	5/5 (100%)
Peer Review (RCR-SBE)	10/13/13	5/5 (100%)
Conflicts of Interest (RCR-SBE)	10/13/13	6/6 (100%)
Collaborative Research (RCR-SBE)	10/13/13	5/5 (100%)
The CITI RCR Course Completion Page	10/13/13	No Quiz
University of Delaware	10/13/13	5/5 (100%)

For this Completion Report to be valid, the learner listed above must be affiliated with a CITI Program participating institution or be a paid independent Learner. Faisified information and unauthorized use of the CITI Program course site is unethical, and may be considered research misconduct by your institution.

Paul Braunschweiger Ph.D. Professor, University of Miami Director Office of Research Education CITI Program Course Coordinator