# **EVALUATING THE QUALITY**

# OF EARLY CHILDHOOD EDUCATION PROGRAMS

By

Dale C. Farran, PhD

Kerry Hofer, PhD

Peabody Research Institute

VANDERBILT UNIVERSITY

Nashville, TN

February 2011

In O. Saracho & B. Spodek (Eds.), *Handbook of Research on the Education of Young Children*. New York, NY: Routledge/Taylor & Francis.

## ABSTRACT

In this paper we review efforts to find measures that can evaluate the quality of early childhood education, an effort that has been complicated by the very different histories and missions of programs in this field. We split the types of early childhood education programs in the United States into two types: ones with a caring mission (child care) and those with a compensatory education mission (prekindergarten and Head Start). The original measures of quality developed to assess child care environments helped to bring attention to the needs of young children and made states aware of their responsibilities to do more than just make sure these locations were healthy and safe. These quality measures have not, however, transferred effectively to programs with a compensatory education mission. For those programs, the clear implication of their mission is that children's academic outcomes must be improved. The academic outcomes found to be most important for later school success are math, reading (language/literacy) and attention. As we show here, none of the quality measures currently in the field have demonstrated much capacity for identifying classrooms that are more effective in helping children learn those skills. We argue that the first step in developing an effective measure of classroom quality has got to be empirical investigations of the behaviors of teachers and children demonstrated to be linked to gains in those three skill areas.

# EVALUATING THE QUALITY OF EARLY CHILDHOOD PROGRAMS Introduction

In 2008, 67% of the four year old children in the United States were enrolled in some form of early childhood education center-based program (Planty, Hussar, Snyder, et al., 2008), and the number continues to rise. Today's preschoolers will soon begin their formal schooling years. In another 15 years, most of those children will be entering the country's work force. The recent extension of public education into the prekindergarten years make early childhood settings the first introduction for many children to the world of more formal learning and to doing so in a group setting. These early experiences are critical for establishing learning and dispositional patterns that may affect children's interactions with classrooms for years to come. Over the years, however, no clear or coherent focus has emerged for the purpose of early childhood education nor whether there should be different purposes in caring for or educating young children.

Historically, the purpose of caring for young children was to allow impoverished mothers to work: "Day care was founded, therefore, as a necessary social service to alleviate the child care problems of parents who had to work and to prevent young children from wandering the street" (Scarr & Weinberg, 1986, pg. 1141). By the 1970s, it was no longer just immigrant or seamen's families who needed care; mothers were entering the labor force to preserve their families' middle level incomes, and they also needed care for their children. Concurrently with this surge of mothers entering the labor force was a growing concern for the quality of the alternative care the children received. Scarr and Weinberg described this history in detail through the mid 1980s. They report that in 1980 a group of professionals from different agencies and universities developed and proposed a national set of standards for child care that would

cover educational aspects as well as the health and safety issues states were ordinarily concerned with. These standards were never adopted, and since then there has not been a uniform, agreed upon set of standards for the care of children before formal schooling. While all states have regulations concerning health and safety and these sometimes also include educational requirements for teachers and regulations about teacher-child ratios and group size, about half the states have developed their own more comprehensive quality standards. In July 2010, the National Association for the Education of Young Children (NAEYC) reported that 24 states had *statewide* quality rating and improvement systems (NAEYC, 2010).

Another role for programs for young children prior to formal school entry has been as compensatory education, beginning with Head Start in 1965 (Farran, 2007; Scarr & Weinberg, 1986) and continuing with the 1987 amendment to the Elementary and Secondary Education Act that allowed Title I funds to be used for whole school program improvement, ushering in the creation of Title I funded prekindergarten classes in many school districts (Ewing & Matthews, 2005). Neither Head Start nor Title I was intended to be full day care; the usual hours of care for each have been public school hours or less that do not cover before or after school care needed by working families. Although some programs provide these services as an option, many do not. Recently, there has been an increase in the number of children served in states that either provide state funds for early intervention prekindergarten programs or that coordinate sources of funding for these programs at the state level. In 2009, 38 states funded preschools and enrolled over one million four year olds (Barnett, Epstein, Friedman, Sansanelli, & Hustedt, 2009); this is not much more than the number of states who provided such funds in 1998 (Mitchell, Ripple, & Chanana, 1998), but those states with programs enrolled more children each year (until this past

one). These state-funded programs are primarily intended as compensatory for children from poor families; 32 of the states have income requirements for enrollment.

Given the different missions of community child care and compensatory preschools, it has been difficult to find a common quality measure that was suitable for both types of program. The importance of aligning quality measures with the purposes a program is to achieve is explored in the next section.

## **Definitions of Quality by Programmatic Purpose**

## **Caring Purpose**

The purpose of community child care programs is to care for children whose parent work. The concerns related to this purpose revolve around ensuring safe and appropriate environments for children who are young and vulnerable and therefore dependent on adults to create their environments. Moreover, when other adults, outside of "kith and kin," are responsible for children they may be unknown to the parents, at least initially. Thus, regulatory agencies become involved. Child care centers are supposed to be licensed to meet health and safety regulations and are inspected by those agencies<sup>1</sup>. In addition increasingly over the past 15 years, childcare programs across the country have been evaluated for their quality, using standardized rating forms administered in various ways, but most often during a visit from the Day Care division of Departments of Human Services. Scores presented as indications of the quality of a program are often made publicly available to parents in order to aid them in the selection of their child's preschool placement. Indiana's *Paths to Quality* is a good example of a Quality Rating System; run by the Indiana Family and Social Services Administration, its web site

<sup>&</sup>lt;sup>1</sup> Unregulated child care programs exist in all states, as a type of underground economy, and may avoid even the basic licensing requirements.

(<u>http://www.in.gov/fssa/2554.htm</u>) allows parents to find ratings of child care centers they are considering.

These quality scores are also used by the policy world in decisions about program funding and childcare reform. Currently many states assess child care quality to determine the amount of money that is awarded to preschool programs in state support for the care of children from low income families. This process effectively ties a program's score to, among other things, the salary of the program's staff. This type of real consequence in the use of a quality measure lends urgency to a determination of the validity of such measures.

## **Education Purpose**

Starting from quite a different point and perspective, programs whose primary purpose is compensatory education may be very separate from programs for whom caring is primary. Head Start programs do not typically provide full-day or year round care and are therefore not good sources of care for working parents<sup>2</sup>, but most Head Start programs are licensed and reviewed in ways similar to child care in their respective states. Prekindergarten programs funded by Title I or by state funds, on the other hand, are often located in public school buildings or connected to Local Education Agencies (LEAs); this connection, for example, is required for programs receiving state funding in Tennessee. Historically these programs have been resistant to coming

<sup>&</sup>lt;sup>2</sup> While some Head Start programs arrange full day options for working parents, they are not required to by law. The most recent Head Start Program Standards to be found online (<u>http://edocket.access.gpo.gov/cfr\_2007/octqtr/45cfr1306.32.htm</u>) list the following as the hours of care:

<sup>&</sup>quot;(b) Center-based program option requirements. (1) Classes must operate for four or five days per week or some combination of four and five days per week. (2) Classes must operate for a minimum of three and one-half to a maximum of six hours per day with four hours being optimal." Moreover, programs do not operate year round: "Programs that operate for four days per week must provide at least 128 days per year of planned class operations. Programs that operate for five days per week must provide at least 160 days per year of planned class operations"

under the supervision of the Departments of Health and Human Services and day care regulations.

In order to create a standard for evaluating these types of state-funded programs, the National Institute for Early Education Research (NIEER) developed a set of 10 Benchmarks on which states are graded each year. None of the benchmarks requires actual observations of the classrooms. Instead they deal with regulatory issues such as the adoption of early learning standards in the state, the requirement that lead teachers in each classroom have a bachelor's degree, the condition that assistants have a Child Development Associates degree (or equivalent), as well as issues of staff-child ratios and group size. NIEER gathers the data related to the Benchmarks each year and issues a yearbook giving state by state evaluations (Barnett et al., 2009).

## **Outcomes Related to Differences in Programmatic Goals**

Measures of quality logically should be related to the types of outcomes desired. While it might be the case that a universal measure of quality could encompass the outcomes intended by both types of programs described above, it is unlikely. In the general literature, quality is described in terms relating to excellence, value, conformance to specifications, and/or meeting customer expectations (Reeves & Bednar, 1994). Since the expectations for these two types of programs are quite different, it has been difficult to find a single measure to capture the quality of both. It is important to note that this paper is focused on the expectations of two types of early childhood programs in the United States; other countries have different expectations about young children's development, have different programmatic structures, and therefore different quality standards.

For the caring mission, the emphasis has been more on *preventing* the presumed deleterious consequences of poor quality care. By assuring that child care environments are safe, organized, material-rich and filled with positive teacher child interactions, children should develop typically. The types of outcomes often examined as a function of *poor* quality care are socio-emotional in nature (to be discussed in detail later). This perspective has been influential on the types of quality measures developed for these environments. Group care for young children in the U.S. is stressful as recent studies by Gunnar and her colleagues have shown (Gunnar, Tout, de Haan, Pierce, Stansbury, 1997; Watamura, Kryzer, & Robertson, 2009). The evidence is that children evinced greater stress as the day progressed even in high quality child care centers, though quality was found to ameliorate the effect.

For the education mission, much greater clarity would be expected for the definition of outcomes to be obtained. In the U.S., these programs were specifically established to prevent school failure for children from poor families. Determining which proximal outcomes at the end of pre-k will be associated with long term school success is therefore important. One source of useful guidance is the recent analysis of data from six major longitudinal studies from the United States, Canada and Great Britain (Duncan, Dowsett, Claessens et al., 2007), findings confirmed by analyses from additional longitudinal studies (Hooper, Roberts, Sideris, Burchinal, & Zeisel, 2010). Gains in reading (or language and literacy at pre-K) during the pre-K and kindergarten years were found to be associated with better performance in reading at grades 3 to 5, a finding consistent with other longitudinal studies of reading (Schatschneider, Carlson, Francis, Foorman, & Fletcher, 2002). Also early measures of children's math skills predicted later reading and math. Indeed, early measures of math were somewhat more predictive of later reading achievement than were early measures of reading skills. Finally, Duncan et al. found measures of

cognitive self regulation (distinguished from emotional regulation and social skills) to be a significant correlate of later achievement. Cognitive self regulation consists of effortful control of attention, task persistence, sustaining attention and inhibition of impulses.

Both Duncan et al. and Hooper et al. concluded that the best skills to effect in these prekindergarten programs are those related to math, reading, and attention, in that order. Knowing these should be the desired outcomes for education-focused prekindergarten programs provides a lens through which to judge measures of quality; the utility of a quality measure for these classrooms should be judged by its association with gains for children in these three areas.

## **Caring Mission: Measurements and Consequences of Quality**

## Measures

There are a number of instruments designed to assess general quality in early child care environments, and several of these instruments have been shown to measure similar aspects of the classroom. These instruments differ in their concentration either on the classroom environment as a whole, including physical characteristics, or on instruction and interactions alone. Though each has a different focus, many of the instruments are consistently used in the assessment of early childhood classrooms both domestically and internationally. One instrument designed to assess quality is the Early Childhood Environment Rating Scale (ECERS, Harms & Clifford, 1980), it is perhaps the most widely-used measure to evaluate program quality (Sakai, Whitebook, Wishard, & Howes, 2003).

Because of its extensive use, the ECERS has become an anchor scale for other instruments. Many research studies using other instruments to assess quality correlate those instruments with the ECERS. Scores from the ECERS have been correlated with the Classroom Assessment Scoring System (CLASS; LaParo & Pianta, 2003) and the Snapshot (Ritchie,

Howes, Kraft-Sayre, & Weiser, 2002) in reports from the National Center for Early Development and Learning's Multi-State Pre-Kindergarten Study (LaParo, Pianta, & Stuhlman, 2004; Pianta et al., 2005). Additionally, the ECERS has been correlated with the Caregiver Interaction Scale (CIS; Arnett, 1989), the Early Childhood Observation Form (ECOF; Stipek, Daniels, Galuzzo, & Milburn, 1992), and the Adult Involvement Scale (AIS; Howes & Stewart, 1987). Researchers have also found correlations between the ECERS and the Assessment Profile for Early Childhood Programs (Abbott-Shim & Sibley, 1987), a similar measure examining the quality of the classroom setting (Phillips, Mekos, Scarr, McCartney, & Abbott-Shim, 2000; Scarr, Eisenberg, & Deater-Deckard, 1994). Although some of these instruments relate more strongly to the ECERS than others, researchers have found statistically significant correlations between the ECERS (total score and/or subscale scores) and all of the above-mentioned measures (total score and/or subscale scores), some as high as r = .91.

## Socio-emotional Outcomes

It is difficult to address socio-emotional development in relation to quality during the preschool years. Measures of socio-emotional development are usually obtained through adult ratings, provided either by teachers or parents, each of which has its problems as will be described in this section. The better studies in this area obtain the teacher ratings when children are in kindergarten or beyond by individuals not associated with the quality of the child care program. Positive findings from these sorts of studies would go a long way to validate the use of a measure like the ECERS for programs whose mission is in the care domain. There have been a number of such studies, and the findings are mixed.

Two studies followed children into elementary school whose preschools had been rated previously with the ECERS. In a study from the United Kingdom, Sylva, Siraj-Blatchford, Taggart, et al. (2006) found ECERS scores to be related to first grade teacher ratings of

#### EVALUATING QUALITY IN ECE

children's cooperation and conformity. Higher scores on the ECERS total score as well as the Interaction scale in particular were associated with more positive social outcomes later. On the other hand, Peisner-Feinberg, Burchinal, Clifford et al. (2001), using a composite score of quality that included the ECERS total, found no relationship between quality and second grade teachers ratings of children's social competence. Both of these studies were focused on a heterogeneous population of children in community child care programs. It is conceivable that the quality of child care might be more important for children who come from high risk circumstances as the following studies examined.

The "Three-City Study" has followed a large sample of children from low-income families whose mothers were required to return to work following welfare reform in 1996 (Loeb, Fuller, Kagan, & Carrol, 2004; Votruba-Drzal, Coley, & Chase-Lansdale, 2004; Votruba-Drzal, Coley, Maldonado-Carreño, Li-Grining, & Chase-Lansdale, 2010). The child care arrangements of the children were documented; child care centers were observed, and a quality score was determined from a combination of the ECERS and the Caregiver Interaction Scale (CIS). The initial study examined socio-emotional competency in the children as preschoolers from ratings given by their mothers. Higher CIS scores were associated with fewer ratings of social problems by the mothers, while the ECERS total score was not (possibly because the ECERS scores in the three cities differed significantly from each other) (Loeb et al., 2004). In the same study, Votruba-Drzal et al. (2010) found an interaction between reports of problem behaviors, the ECERS total score and time in care – the more hours a day children were in low quality care, the higher their externalizing scores. The importance of global child care quality (that includes the teacher-child relationship) emerged when they followed the children into middle grades: "The most consistent results from this analysis highlight the importance of quality of care in the reduction of problem behaviors" (Votruba-Drzal et al., 2010, p. 1469). The ability of high quality child care to reduce the emergence of behavioral problems later in children of poor families was particularly apparent for boys and for African-Americans.

Thus it appears that a global quality rating like the ECERS, perhaps best in combination with a more in depth measure of the teacher's interaction style with children, is associated with the kinds of outcomes with which care-mission programs are most concerned – the long term social and emotional outcomes of early care. While child care quality appears in some studies to be an important factor for all children, the studies reviewed indicated that quality of early care rated generally and globally was especially important for males and African-Americans from low income families. Assuming that programs with a mission of caring for young children should be most concerned with social and emotional long term outcomes, the ECERS and other more focused interaction measures appear to document aspects of quality related to the purpose of the programs.

#### Academic and Cognitive Outcomes

The predictive utility of these global measures has been less successful in predicting cognitive outcomes for children. While one could argue that it was not the intention of programs with a caring mission to effect change in academic outcomes, measures like the ECERS have been used for just this sort of investigation. A primary difficulty with much of this research for determining the usefulness of this quality measure to track academic outcomes is the fact that many of the studies did not use pre and post measures and could not assess *gain across time* in relation to quality. As would be expected, there is a strong relationship between child care quality scores and children's skill levels, given that parents are the ones who choose and must pay for the programs. Parents with higher incomes and more education have many more choices and the freedom to choose higher quality programs; their children are also likely to be better prepared for school. A few studies have examined change over time and related the amount of change in children's skills to measures of child care quality.

For example, in a longitudinal study exploring the effects of preschool, Sylva et al. (2006) examined 26 preschools in England for the relationship between child care environmental quality and children's development. She and her colleagues used both the ECERS and the ECERS-Extension (ECERS-E; Sylva, Siraj-Blatchford, & Taggart, 2003). The ECERS-E was developed as a supplement to the ECERS. Developers of the extension argued that the ECERS did not devote enough attention to the cognitive and pedagogical demands of the classroom necessary for children's intellectual and social development (Sylva et al., 2006). Three of the ECERS-E's four subscales refer to a specific academic environment (literacy, science, and math), and the fourth subscale examines the emphasis on diversity within the classroom.

On average, participating programs scored in the adequate to good range on the ECERS and in the adequate range on the ECERS-E. Children were assessed for their cognitive skills and language knowledge at age 3 and again at age 5 using the British Ability scales. At age 5, children were also given a test of letter recognition and phonological awareness. After controlling for age, pretest scores, and child and family background variables, quality as measured by the ECERS-E (both total and subscale scores) was significantly predictive of children's post-test scores on the pre-reading, general math concepts, and non-verbal reasoning skills assessed at age 5. Effect sizes, however, were fairly small, ranging from .11 to .17, indicating that the average difference in child scores due to increases in classroom quality was less than one fifth of a standard deviation. To illustrate what this effect size means from data in this study, the effect of average total ECERS-E scores on children's gains across the year in prereading skills was reported as approximately .17. The authors reported a raw grand mean and standard deviation for this outcome as 21.57 and 12.67, respectively, indicating fairly strong variability in this outcome across classrooms. An effect size of .17 indicates that differences in the pre-reading gains based on classroom quality (measured by the Extension of the ECERS) were likely around 2.15 points. However, the article did not explicitly state their results this way and, as such, it is difficult to interpret their original effect sizes. Suffice it to say, they were modest. For the other outcomes measured, ECERS-E quality scores did not predict gains in spatial awareness or language.

On the other hand, the more commonly used quality measure, the total ECERS quality

score, was not significantly related to gains in any of the child academic outcomes at age 5 although one of the subscales, Interaction, was related to gains in children's general math concepts scores at age 5 (with an effect size of .199).

Two studies of academic outcomes in connection to global child care quality have results that seem both conflicting and hard to resolve, one reporting a significant effect on math outcomes but not reading outcomes, and the other reporting exactly the opposite. As part of a large-scale study of center-based child care and longitudinal child outcomes, the Cost, Quality, and Child Outcomes in Child Care Centers Study, researchers assessed program quality and child outcomes in childcare, kindergarten, and second grade (Peisner-Feinberg et al., 2001). Preschool quality was measured with a combination of instruments, the ECERS, CIS, ECOF, and AIS, and the scores were combined through a principal components analysis to yield one composite quality index for each classroom. Children's receptive vocabulary, letter-word knowledge, and pre-math abilities were assessed in kindergarten and second grade. The quality of children's elementary school classrooms in kindergarten and second grade was also rated. Hierarchical regression analyses revealed that, when maternal education, ethnicity, gender, age, quality in kindergarten and second grade, and teacher-child relationship measures were included in the model, only math outcomes were significantly predicted by preschool quality, a similar finding to that of Sylva et al. (2006). Children from childcare environments of higher quality tended to have higher math outcomes in second grade. Researchers did not find a predictive relationship between childcare quality and vocabulary or letter-word outcomes.

In contrast to the findings of Peisner-Feinberg et al. (2001), a similar study examining the effects of child care experiences through sixth grade did not find a relationship between quality and math outcomes (Belsky et al., 2007). This study was part of the National Institute of Child Health and Human Development Study of Early Child Care and Youth Development that began in the 1990s and examined the effect of child care on longitudinal outcomes, following children from birth through sixth grade. Researchers observed the quality of childcare that children experienced at 6, 15, 24, 36, and 54 months of age using the ORCE, a measure of the quality of

#### EVALUATING QUALITY IN ECE

the caregiver-target child interactions. During the formal elementary years, researchers used the COS to measure quality in first through fifth grade classrooms that included target children. Both the ORCE and the COS differ from the ECERS and other global classroom measures in that each focuses on the individual target children and the environment around those children. Children's outcomes were assessed in preschool, first, third, and fifth grade, with either a letterword knowledge task or general reading task (based on age), an applied problems mathematics assessment, and an assessment of expressive vocabulary knowledge. Controlling for child and family demographic measures, the only fifth-grade outcome that was significantly predicted by child care quality was expressive vocabulary. Children who experienced higher quality child care had slightly higher expressive vocabulary scores in fifth-grade. This relationship was not found for the reading measure or the math measure; reading scores were predicted by previous child care quality through kindergarten but not beyond.

Upon first glance, it seems confusing, especially for determining a good quality measure for child care, that two large longitudinal studies of child care quality and academic outcomes could report such contradictory findings. However, the two studies and the British study of Sylva and her colleagues share one common characteristic; none was experimental. Rather, all studies tracked children through their early years, measuring characteristics of their home and care experiences as they transitioned to and progressed through school. All three of these studies are important; child care is not an experience that lends itself easily to random assignment. When random assignment and group design are not utilized, however, differences in study groups, both before and after attrition, and differences in the experiences of those participants in study classrooms can be quite great. Descriptive studies of this type have been more common than experimental ones in the area of child care because in early childhood research it is really impossible to experiment with variations in the quality of care young children receive. This is not the case for programs with an education mission, as the next section will demonstrate.

## **Education Mission: Measurements and Consequences of Quality**

#### Measures

There are both more measures and more types of measures associated with assessing the quality of programs whose mission is compensatory education than there were with child care. The ECERS, however, has been used extensively to measure the quality of educationallyoriented programs, especially those connected with Head Start, even though its roots are in child care. We have mentioned the efforts in England to develop a supplement to the ECERS that is focused specifically on the academic environment, the ECERS-E. Sylva and other international researchers are the only ones so far to have done much work with that measure, though it appears promising. The Early Language and Literacy Classroom Observation (ELLCO, Smith, Dickinson, Sangeorge, & Anastasopoulos, 2003) is a recent development that has an ECERS type format for its ratings together with a checklist for particular environmental features associated with literacy instruction. So far, it appears to have been used primarily for professional development, as an aid to help teachers set up their prekindergarten classrooms to focus on literacy skills. An instrument rapidly increasing in use is the Classroom Assessment Scoring System (CLASS, Pianta, La Paro, & Hamre, 2008). This rating system is structured differently from the ECERS; observers watch interactions within the classroom for a period of time, usually 20-30 minutes and then give 1-7 ratings on 9 scales grouped into 2 or 3 dimensions. CLASS ratings can also be done from videotapes. In either case, total observation time can range from 2 hours to all day; in some studies observers visited the classrooms twice and averaged the ratings across the days. As with the ECERS, the focus of the observation is on the classroom as a whole, and the CLASS is focused particularly on the behavior of the teacher.

Other systems focus on individual children in the classroom. A picture of the classroom as a whole can be determined by aggregating scores from the individual children; alternatively, analyses are conducted with children's scores nested within classrooms. Three such measures are of note. The first is the *Emerging Academics Snapshot* (Ritchie et al., 2002), sometimes just referred to as "Snapshot," the second is an eco-behavioral system developed by Kontos and Keyes (1999), and the third is a measure developed by Farran and colleagues specifically for use in prekindergarten classrooms, the *Child Observation in Preschool* (COP) (Farran, & Son-Yarbrough, 2001; Farran, Son-Yarbrough, Silveri, & Culp, 1993)<sup>3</sup>. Although not much research has been done linking child focused systems for examining quality with those focused on the teacher/whole classroom, they appear to be measuring different and independent attributes of quality in the classroom.

In the following sections, we will examine the effects of variations in quality as measured by one or more of these systems on the outcomes Duncan et al. (2007) and Hooper et al. (2010) demonstrated are important for prekindergarten programs focused on compensatory education: literacy/language, math, and attention. First, however, we will present information on attempts to link the NIEER Benchmarks to outcomes. These structural features related to regulations are quite different from attempts to capture within-classroom variability in quality. Structural features have featured prominently in requirements for programs with a child care mission, though we did not review their effects in that section. These features are very important for programs with a compensatory education mission because they are key components of the NIEER Benchmarks being used as the guidelines for states establishing prekindergarten programs for children from low income families and because they have recently been questioned in several large studies.

## Structural Quality and Outcomes

<sup>&</sup>lt;sup>3</sup> While the M-ORCE and its originator the ORCE are focused on individual children in the classroom, they have been used almost exclusively in studies of child care and will not be reviewed in this section.

The most ambitious and comprehensive study of the relationship between structural characteristics and child outcomes can be found in the National Center for Early Development and Learning's Multi-State Study of Pre-Kindergarten (Multi-State) and the State-Wide Early Education Programs Study (SWEEP) (Mashburn, Pianta, Hamre et al., 2008.) These large studies have been rich sources of information about the effects of prekindergarten studies and will be referred to again in this paper.

Although the NIEER Benchmarks are intended to rate overall state policies across 10 dimensions, Mashburn et al. created individual measures for the 671 classrooms included in the study. Because these are state-funded prekindergarten programs, many of the benchmarks were met by 70% or more of the programs; for example, most tended to have teachers with BA degrees, to have class sizes of 20 or under, and to have student-teacher ratios of 10:1. They served meals and provided family supports (though the latter is not always specified in detail). They were somewhat less likely to use a comprehensive curriculum, offer health services, or require their Assistants to have a CDA. It is important to list these details because it is variation in the latter three that is likely to carry the weight of prediction for outcomes, given the smaller amount of variation in the other benchmarks.

Associations between gains in language, literacy and math measures and the benchmarks, collectively and individually, were examined. The total NIEER benchmark score was related to none of the outcomes. Similarly none of the individual benchmarks predicted gains on the outcomes with the exception of a negative relationship between serving meals and gains in receptive vocabulary, likely a function of the fact that programs serving poorer families are required to serve meals.

The relationship between teacher credentials and child outcomes has been the focus of intense study over the past few years. The same Multi-State study provided more detailed data on teacher preparation (years of teaching, highest degree earned, whether the teacher had a bachelor's degree or licensure) and found no effects on children's gains in language and literacy for any of the teacher preparation measures (Early, Bryant, Pianta et al., 2006). The only outcome for which teachers' education had a positive effect was a small one on math skills. Exploring the issue of teacher credentialing further and in samples with much greater variation in teachers' educational status is the recent analysis of data from seven major studies, the majority of which involved prekindergarten or Head Start programs (Early, Maxwell, Burchinal, Alva, Bender, Bryant, et al., 2007). The overwhelming conclusion across these seven studies is that neither the presence or absence of a BA degree, having a teaching license, or majoring in Early Childhood Education were related to children's gains on either language/literacy or math measures.

The surprising lack of relationship between structural measures of quality, especially teacher credentials, requires careful reflection. A belief in the importance of both structural characteristics and teacher credentials has been strong at least since the publication of *Children at the Center* (Roupp, Travers, Glantz, & Coelen, 1979); the requirement that teachers have a bachelor's degree is the primary recommendation of the National Research Council's 2001 report on preschool education. One possible explanation for the lack of relationship is two pronged – children are not making a great deal of gain in these prekindergarten programs, especially in the areas of language and mathematics (Howes, Burchinal, Pianta, et al., 2008; U.S. Department of Health and Human Services, January 2010), and teachers are not observed delivering very high quality instruction in their classrooms (e.g., Justice, Mashburn, Hamre, &

Pianta, 2008). It appears that having a BA degree is not sufficient to prepare teachers to be effective in classrooms where the purpose is to work specifically with children whose school entry skills are low. Teaching young children who are developmentally quite different from each other has many challenges (see Farran, 2005); working with children who are enrolling in formal education for the first time and who are there because they do not possess school entry skills is the most formidable challenge of all. It is not clear where teachers would have been expected to learn these unusual, multi-dimensional instructional skills, nor has there been sufficient research to determine if these skills could be obtained through professional development activities.

## Language and Literacy Outcomes and Prekindergarten Quality

Data available from several longitudinal studies of prekindergarten quality have found small and mixed effects for quality on language and literacy gains over the prekindergarten year. The most common measure of language is the *Peabody Picture Vocabulary Test* (PPVT), a measure of receptive vocabulary, although the Multi-State Study and SWEEP also used the Oral Expression Scale from the *Oral and Written Language Scale* (OWLS). Gains in the PPVT and OWLS were analyzed in relation to the ECERS total score and to the two CLASS dimensions: Emotional Support and Instructional Support. The findings are presented in several publications from this major study (e.g., Howes et al., 2008; Mashburn et al., 2008). None of the three quality measures predicted gains on the PPVT; the PPVT is a complex measure of language, focusing not just on nouns (names of things) but also on verbs, adjectives and adverbs. The total score on the ECERS was associated with gains in expressive language measured by OWLS, but the effects were modest. ECERS total scores and the two CLASS dimensions were also examined for their effects on gains in literacy skills (e.g., letter naming, rhyming, print awareness) (Howes et al.; Guo, Piasta, Justice, & Kaderavek, 2010). There were similar, small effects (*d* = .06 and

.07) on these outcomes. Overall, it appears that the gains were small in these programs, although there was classroom variation in how much children gained. However, it appears that the summary ratings from ECERS or CLASS are accounting for very little of the variation in children's developmental outcomes.

The CLASS observations have been explored in a different way. Rather than using summary scores (a variable analysis), LoCasale-Crouch created profiles of teachers from scores on the CLASS dimensions and item scores within each dimension (LoCasale, Konold, Pianta, et al., 2007). Profile 1 was labeled the "highest quality" – the teachers in this cluster had the highest CLASS scores on Emotional Support; these classrooms also had the highest ECERS scores. The profile with the lowest CLASS scores across all the dimensions was Profile 5. Profile 5 also had the lowest ECERS scores, included the most Head Start classrooms, and enrolled the highest proportion of poor and minority children.

In a follow up study, Curby and colleagues examined the fall and spring academic gains for children who were taught by teachers in each of the profiles (Curby, LoCasale-Crouch, Konold, et al., 2009). Children taught by Profile 2 teachers made substantially greater gains on the PPVT, the language measure, than children taught by teachers in Profile 5, with an effect size of d = .21. Teachers in Profile 2 were not as warm or emotionally supportive as teachers in Profile 1, but they had the highest scores on the CLASS item, Concept Development, of any of the profile clusters. Concept Development is one of the items scored under Instructional Support on the CLASS. No data on early literacy assessments were presented as outcomes even though these two reports are focused on findings from the Multi-State Study and SWEEP.

In sum, examining the quality of prekindergarten programs for its effects on language/literacy outcomes leads to the conclusion that there is a lot of work to do to find an

instrument that will capture important differences among these classrooms that relate to child gain. Curby et al. report substantial classroom variance in the outcomes studied but it appears that little of that variance is related to current measures of quality.

An alternative potential measure of quality and its relation to language and literacy involves focusing on child behavior rather than the teacher's. Most of the research in this area is descriptive, and few studies have tried to link summaries of child behavior to prekindergarten child outcomes. The Multi-State Study and SWEEP collected Snapshot, CLASS, and ECERS data from the same classrooms in which they assessed children's behaviors. These large studies only focused on four children per classroom; children were observed from the beginning of class to naptime on two different days. Chien, Howes, Burchinal, et al. (2010) used Snapshot data to create profiles of children's time use across the day. Profile 1 was characterized as the "free play" profile and included 51% of the children; these children spent less time in pre-academic activities than children in any of the other three profiles. Profile 2, constituting only 9% of the children observed, was labeled "individual instruction." Profile 2 children spent more time on activities like worksheets or computers, activities that involved individual work. Profile 3 (27%) spent more time in whole and small group instruction, while the 13% of the children in Profile 4 spent their time in scaffolded instruction.

These profiles of experience in prekindergarten classrooms were associated with gains in language and literacy measures in surprising ways. There were no differences in language outcomes among the four profiles for more complex measures of language like the PPVT and OWLS. On letter naming and writing legibly, children in Profile 2 made significantly more gains than children who had experiences associated with the other profiles. What is important is that the Profile 2 children from poor families made the most gains of any group including non-

poor children in the other profiles. While these effects are very small (d = .04), the idea that we might capture the quality of a prekindergarten classroom by examining the profile of experiences children have during the day is intriguing. Of interest also is the fact that classroom environmental characteristics associated with this profile were not captured by the ECERS scores. In fact, the highest ECERS scores were found in classrooms characterized by Profile 1, which contained children who showed less growth in all measures compared to children in the other profiles.

In sum, efforts to link existing quality measures with gains in complex language skills have been unproductive for the most part. Children do not make much gain in these measures, and the gains they do make seem to be either unrelated or very modestly related to the currently extensively used quality measures of prekindergarten environments. The Snapshot approach appears promising, but it, too, only showed relationships to gains on the more concrete measures of literacy such as learning the names of letters and writing one's name.

## Mathematics Skills and Prekindergarten Quality

Despite the fact that Duncan et al., 2007 and Hooper et al., 2010 listed mathematics skills as the first set of skills important to effect in prekindergarten programs (followed by reading and attention), much less attention has been paid to what facilitates the development of skills in this area for young children from poor environments. In fact, unlike language and literacy, one primary test of skill development dominates the research that has been done in early mathematics, the Woodcock Johnson III Applied Problems Subscale. Several of the studies already reviewed also included Applied Problems as a child measure; no study examined growth in mathematics skills in depth and on as the sole focus of the study. Growth in mathematics was predicted at about the same small magnitude of effect and by many of the same quality measures as the literacy measures. Ratings on the CLASS Instructional Support subscale predicted gains in Applied Problems (Mashburn et al., 2008), and, it follows, that children who had teachers in Profile 2, the profile with the highest Concept Development scores, made more gains in mathematics. The effect size of 0.19 on Applied Problems between children who had Profile 2 teachers and those whose teachers were Profile 5 is one of the largest obtained in this set of studies. Finally, children who displayed the Individual Instruction profile (Profile 2) also showed more gains in math as well as the literacy measures.

Another investigation of child behavior in prekindergarten and Head Start classrooms used the *Child Observation in Preschool* (COP) to measure such child behaviors as verbalizations, engagement, and time spent with various materials (Cummings, Hofer, Farran, & Lipsey, 2010). The COP is organized so that researchers can examine contingent probabilities, one of which involves children verbalizing while engaged in a math activity. In this large study of an early mathematics curriculum enacted in prekindergarten and Head Start classrooms, talking while engaged in math was relatively rare, but significantly associated with gains in mathematics across the prekindergarten year.

In sum, relatively less is known about the relationship between quality measures and mathematics development for children in prekindergarten classrooms. Promising areas for further study involve the complexity of teacher instruction, time to practice mathematics skills in individual work, and the encouragement of child verbalization when engaged with math tasks. It does seem as if these three areas could be straightforwardly incorporated into a measure of quality that includes the importance of mathematics gains among its foci

## Attention Skills and Prekindergarten Quality

The quality of prekindergarten classrooms associated with gains in attention has been characterized by neglect. Attention is a positive child behavior that belongs in a cluster of behaviors often called "learning dispositions" (Katz, 1999). One of the great difficulties in studying this important area is the lack of validated measures, especially any that are not based on teacher ratings. Teacher ratings of learning dispositions taken during the prekindergarten year are difficult to use as an outcome measure against which to investigate the effects of quality. This is especially the case if teachers are aware of the focus on dispositions and would like their children to make improvements. One of the few studies to focus on attention is the Dominguez, Vitiello, Maier, and Greenfield (2010) study of 275 children in 29 Head Start classrooms. Teachers rated children's "learning behaviors" three times over the prekindergarten year. Their classrooms were observed with CLASS once during the year.

There was considerable variation among the classrooms in the rate of change in learning behaviors portrayed by the teachers. Organization was the only CLASS item related to these ratings, accounting for 1% of the variation in the rates of change in learning behavior. The lack of relationship between classroom quality measures and ratings of learning behaviors led the researchers to conclude that "additional classroom-level predictors should be examined" (p. 42). Social competencies and problem behaviors have more often been investigated although these are not the behaviors so far shown to have long term significance for school achievement. Social competence and problem behaviors are not measures of attention; in fact, Duncan et al. separated social skills ratings from attention ratings and found that social ratings accounted for very little later of the variation in children's school achievement while, as already noted, early measures of attention were significant and independent predictors. The large multi-state studies of the National Center for Early Development and Learning included measures of social competence

and problem behaviors in the children, with ratings provided both by the prekindergarten teachers and, for a substantial number of children, their kindergarten teachers. During the prekindergarten year, changes in teacher ratings of children's social skills were predicted to a small degree by CLASS ratings of Emotional Support in the classroom (Mashburn et al., 2008); at the end of kindergarten the quality of children's prekindergarten experiences as captured by CLASS had little to no effect on teacher ratings (Curby et all, 2009). In fact, the strongest predictor of social skills was being female, and for having problem behaviors, being a member of a minority group.

The awareness of the importance of attention or learning dispositions is relatively recent, and little research has so far been published to investigate the quality of classrooms that might facilitate its development. But one can imagine that very different aspects of the environment could be involved in helping children learn to focus, persist, and pay attention than would be involved in learning other skills. Clever and careful thinking will be required.

# Issues in Measuring Quality for Both the Care and Education Missions

## The Use of Observer Ratings

The primary measures of the quality of child care and prekindergarten classrooms rely on observer ratings. For the following reasons the use of ratings may, in fact, contribute to the fact that quality has been such an ephemeral construct.

First, ratings require a judgment, and judgments have a subjective component. It is exceptionally difficult for raters to adopt enough of a shared perspective to insure that they are all rating the same qualities of teacher behaviors and classroom organization. Because of the difficulty of agreement, many studies reviewed in this chapter count as "agreement" when raters come within one point of each other. This tendency effectively reduces the scale metric; one rater's score of 3, for example, on CLASS and ECERS, could be rated as a 2 or a 4 and be counted as an agreement. But scores of 2 and 4 on both instruments are assumed to have quite different meanings.

Compounding the problem with the way agreement is determined is the fact that in many studies reliability is only established during the training phase using videotapes provided by the scale developers. This practice has at least two flaws: first, training reliability should always be followed up by field reliability across the length of the data collection period to check for observer drift especially in subjective ratings, and second, reliability should always be established in the type of setting in which data will be collected. A videotape is not the same stimulus as a live classroom; if observers are going to collect data in live classrooms, classrooms must be the places where reliability is obtained.

Second, rating scales begin with the assumption that qualities of classrooms can be described on continua from poor to exemplary along prespecified dimensions. Because not enough empirical work has determined which particular behaviors of teachers and aspects of classroom organization are related to child outcomes of interest, rating scales have emerged from ideological beliefs in the importance of particular qualities. The ECERS reflects a perspective that the materials in the classroom, the ways they are organized and the amount of time children are allowed to explore them freely are critical quality features. The empirical work to determine which aspects of organization, which and how many types of materials, and how to facilitate children's focus during free play has not occurred. CLASS proceeds from a perspective that the emotional atmosphere of a classroom and the teachers' interactions with children are the critical quality features. The point is not to call these ideological perspectives into question but to demonstrate that these are *beliefs* and not empirically determined measures of quality

demonstrated to link to the outcomes of interest to the programs. Curby et al. (2009) concluded their investigation of quality in prekindergarten classes by remarking, "Given their [prekindergarten programs] explicit intention of changing students' school readiness and performance trajectories, it is important to identify classroom practices that promote student learning and, thus could serve as a target for intervention" (p. 364) and, one might add, could serve as the basis for the development of a quality rating scale.

## Child Perspective Versus a Teacher/Classroom Perspective

Most of the current measures of quality focus on the classroom as a whole, in effect, investigating the classroom from the "top down." A "bottom up" or child perspective might provide a very different picture of classroom quality (e.g., Powell, Burchinal, File, & Kontos, 2008). Examining only what the teacher does can be misleading as it does not provide the observer with information about what children are actually receiving. Hofer (2006) conducted a small study of word learning by children during storybook reading. She supplemented pre and post assessments of the children's knowledge of words in the teacher-read stories with observations of child attention during whole group story reading (using the COP). Only the few children who had been rated as attentive during the storybook reading actually learned any of the words. On a measure of the quality of storybook reading, the teachers would have been rated highly. They read the stories well, called attention to the new words, and offered definitions. From a top down perspective, they performed well, but from the children's perspective the lessons were mostly ineffective.

The national evaluation of Early Reading First (ERF) is a good example of the problems in taking a teacher-only perspective (Jackson, McCoy, Pistorino, et. al., 2007). Teachers in classrooms funded by ERF were observed and rated on literacy practices that would seem to reflect the types of instruction that should lead to strong literacy growth in children. The rating scales used were the ECERS and the Teacher Behavior Rating Scale (TBRS; Landry, Crawford, Gunnewig, & Swank, 2004). Teachers in ERF-funded classrooms were rated significantly higher on both ECERS and TBRS than prekindergarten teachers in non-funded classrooms. Despite these large and significant differences in teacher behaviors, no differences were found in child outcomes between those children who had been in ERF classrooms and those who had not. Observing how the children were actually reacting to the increased literacy instruction would have been an informative supplement to these global quality ratings of teacher practices.

A global rating of the classroom by definition blurs individual differences among the children and assumes that the impact of the practices studied will be uniform. The *quality* of a prekindergarten classroom must be concerned with its connection to the needs of the children, even when those needs are quite varied. It may be possible to design a quality measure for early childhood that combines both a classroom and a child perspective, but to date, researchers seem to focus pretty exclusively on one or the other.

#### Summary

Finding a measure that can evaluate the quality of early childhood education has been complicated by the very different histories and missions of programs in this field. The caring mission involves a heterogeneous group of children and families. Children are cared for by adults other than their families away from their homes for large portions of the day. The original measures of quality developed to assess these environments helped to bring attention to the needs of young children and made states aware of their responsibilities to do more than just make sure these locations were healthy and safe. These quality measures have not, however, transferred effectively to programs with a compensatory education mission. For those programs, the clear implication of their mission is that children's academic outcomes must be improved. According to current wisdom, the academic outcomes most important for later school success are math, reading (language/literacy) and attention, and the small number of these skills provides a clear starting point for the development of measures that would be related to growth in these areas. None of the quality measures currently in the field have shown much capacity for identifying classrooms that are more effective in helping children learn those skills. We argue that the first step in developing an effective measure of classroom quality has got to be empirical investigations of the behaviors of teachers and children demonstrated to be linked to gains in those three skill areas.

Recently Hughes (2010) asserted, "The identification of specific classroom transactions or processes that predict the growth in skills that enable children to make a successful transition to kindergarten and first grade is critical to realizing the promise of preschool education" (p.48). Only observational measures that describe specific behaviors (of teachers and children) and examine those behaviors in relation to child growth will be useful in the identification of these important classroom transactions.

## References

- Abbott-Shim, M., & Sibley, A. (1987). Assessment profile for early childhood programs: Preschool, infant and school age. Atlanta, GA: Quality Assist.
- Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10, 541-442.
- Barnett, W.S., Epstein, D., Friedman, A., Sansanelli, R., & Hustedt, J. (2009). *The state of preschool 2009: State preschool yearbook.* Rutgers, NJ: National Institute for Early Education Research, Rutgers Graduate School of Education.
- Belsky, J., Vandell, D. L., Burchinal, M., Clarke-Stewart, K. A., McCartney, K., & Owen, M. T. (2007, March/April). Are there long-term effects of child care? *Child Development*, 78 (2), 681-701.
- Chien, N., Howes, C., Burchinal, M., Pianta, R., Ritchie, S., Bryant, D., Clifford, R., Early, D., & Barbarin, O. (2010). Children's classroom engagement and school readiness gains in prekindergarten. *Child Development*, 81, 1534-1549.
- Curby, T., LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., ... Barbarin, O. (2009). The relations of observed pre-k classroom quality profiles to children's achievement and social competence. *Early Education and Development, 20*, 346-372.
- Cummings, T. P., Hofer, K. G., Farran, D. C., & Lipsey, M. W. (2010). *Changes in preschool children's behaviors mediate the curriculum effect on math achievement*. Manuscript in preparation, Vanderbilt University, Nashville, TN.
- Dominguez, X., Vitiello, V., Maier, M., & Greenfield, D. (2010). A longitudinal examination of young children's learning behavior: Child-level and classroom-level predictors of change throughout the preschool year. *School Psychology Review*, 39, 29-47.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, 43 (6), 1428-1446.
- Early, D., Bryant, D., Pianta, R., Clifford, R., Burchinal, M., Ritchie, S.,...Barbarin, O. (2006). Are teachers' education, major, and credentials related to classroom quality and children's gains in pre-kindergarten? *Early Childhood Research Quarterly*, *21*, 174-195.
- Early, D., Maxwell, K., Burchinal, M., Alva, S., Bender, R., Bryant, D., ... Zill, N. (2007).
  Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development*, 78, 558-580.
- Ewing, D. & Matthews, H. (2005). *The potential of Title I for high quality preschool*. Washington, DC: The Center for Law and Social Policy.
- Farran, D.C. (2005). Developing and implementing preventive intervention programs for children at risk: Poverty as a case in point. In M. Guralnik (ed.), *A developmental systems approach to early intervention: national and international perspectives* (pp. 267-304). Baltimore: Paul Brookes, Publisher.
- Farran, D. C. (2007). Is education the way out of poverty? A reflection on the 40th anniversary of Head Start (with commentaries by James King and Bernard L. Charles), *Monographs* of the Center for Research on Children's Development & Learning, No. 3 (50 pages – ISBN: 0-9727709-2-5).

- Farran, D.C. & Son-Yarbrough, W. (2001). Title I funded preschools as a developmental context for children's play and verbal behaviors. *Early Childhood Research Quarterly*, 16, 245-262.
- Farran, D. C., Son-Yarbrough, W., Silveri, B. & Culp, A. (1993). Measuring the environment in public school preschools for disadvantaged children: What is developmentally appropriate? In S. Reifel (Ed), *Advances in Early Education and Day Care, 1993*, pg. 75-93. Greenwich CN: JAI Press, Inc.
- Gunnar, M.R., Tout, K., de Haan, M., Pierce, S., Stansbury, K. (1997). Temperament, social competence, and adrenocortical activity in preschoolers. *Developmental Psychobiology*, 31, 65–85.
- Guo, Y., Piasta, S., Justice, L., & Kaderavek, J. (2010). Relations among preschool teachers' self-efficacy, classroom quality, and children's language and literacy gains. *Teaching* and Teacher Education, 26, 1094-1103.
- Harms, T., & Clifford, R. M. (1980). *The Early Childhood Environment Rating Scale*. New York: Teachers College Press.
- Hofer, K. G. (2006, October). *Attention and vocabulary learning from shared book reading*. Poster session presented at Vanderbilt University's First Year Research Project Presentations in the Department of Teaching & Learning, Nashville, TN.
- Hooper, S., Roberts, J., Sideris, J., Burchinal, M., & Zeisel, S. (2010). Longitudinal predictors of reading and math trajectories through middle school for African American versus Caucasian students across two samples. *Developmental Psychology*, 46, 1019-1029.
- Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Ready to learn? Children's re-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly*, 23, 27-50.
- Howes, C., & Stewart, P. (1987). Child's play with adults, toys, and peers: an examination of family and child care influences. *Developmental Psychology*, *23*, 423-430.
- Hughes, J. (2010). Identifying quality in preschool education: Progress and challenge. *School Psychology Review*, *39*, 48-53.
- Jackson, R. McCoy, A., Pistorino, C., Wilkinson, A., Burghardt, J., Clark, M., Ross, C., Schochet, P. & Swank, P. (2007). *National Evaluation of Early Reading First: Final Report*, U.S. Department of Education, Institute of Education Sciences, Washington, DC: U.S. Government Printing Office.
- Justice, L., Mashburn, A., Hamre, B., & Pianta, R. (2008). Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early Childhood Research Quarterly*, *23*, 51-68.
- Katz, L. (1999). Another look at what young children should be learning. *ERIC Digest, EDO-PS-*99-5. ERIC Clearinghouse on Elementary and Early Childhood Education.
- Kontos, S. & Keyes, L. (1999). An ecobehavioral analysis of early childhood classrooms. *Early Childhood Research Quarterly*, *14*, 35-50.
- Landry, S., Crawford, A., Gunnewig, S., & Swank, P. (2004). *Teacher Behavior Rating Scale* (*TBRS*). Houston, TX: Center for Improving the Readiness of Children for Learning and Education, unpublished research instrument.

- LaParo, K. M., & Pianta, R. C. (2003). *CLASS: Classroom Assessment Scoring System*. Charlottesville: University of Virginia.
- LaParo, K. M., Pianta, R. C., & Stuhlman, M. (2004, May). The classroom assessment scoring system: findings from the prekindergarten year. *The Elementary School Journal*, 104 (5), 409-426.
- LoCasale-Crouch, J., Konold, T., Pianta, R., Howes, C., Burchinal, M., Bryant, D., ... Barbarin, O. (2007). Observed classroom quality profiles in state-funded pre-kindergarten programs and associations with teacher, program, and classroom characteristics. *Early Childhood Research Quarterly*, 22, 3-17.
- Loeb, S., Fuller, B., Kagan, S., & Carrol, B. (2004). Child care in poor communities: Early learning effects of type, quality, and stability. *Child Development*, *75*, 47-65.
- Mashburn, A., Pianta, R., Hamre, B., Downer, J., Barbarin, O., Bryant, D., ...Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79, 732-749.
- Mitchell, A., Ripple, C., & Chanana, N. (1998). *Prekindergarten programs funded by the states: Essential elements for policy makers.* New York, NY: Families and Work Institute.
- NAEYC (July, 2010). Quality rating and improvement systems (QRIS) and the National Association for the Education of Young Children (NAEYC) accreditation. NAECY Public Policy Fact Sheet, Washington, DC: NAEYC.
- National Research Council (2001). *Eager to learn: Educating our preschoolers*. Committee on Early Childhood Pedagogy. B.T. Bowman, M.S. Donovan, & M.S. Burns (Eds.), *Commission on Behavioral and Social Sciences and Education*. Washington, D.C.: National Academy Press.
- Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., et al. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, 72 (5), 1534-1553.
- 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., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., et al. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science*, 9 (3), 144-159.
- Pianta, R.C., La Paro, K., & Hamre, B. (2008). *Classroom Assessment Scoring System*. Baltimore, MD: Paul Brookes.
- Planty, M., Hussar, W., Snyder, T., Provasnik, S., Kena, G., Dinkes, R., KewalRamani, A., & Kemp, J. (2008). The Condition of Education 2008 (NCES 2008-031). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- Powell, D., Burchinal, M., File, N., & Kontos, S. (2008). An eco-behavioral analysis of children's engagement in urban public school preschool classrooms. *Early Childhood Research Quarterly*, 23, 108-123.
- Reeves, C. A., & Bednar, D. A. (1994, July). Defining quality: Alternatives and implications. *Academy of Management Review*, 19 (3), 419-445.
- Ritchie, S., Howes, C., Kraft-Sayre, M., & Weiser, B. (2002). *Snapshot*. Los Angeles: University of California, Los Angeles.

- Roupp, R., J. Travers, F. Glantz, and C. Coelen. 1979. *Children at the Center: Final Results of the National Day Care Study*. Cambridge, MA: Abt Books.
- Sakai, L. M., Whitebook, M., Wishard, A., & Howes, C. (2003). Evaluating the Early Childhood Environment Rating Scale (ECERS): Assessing differences between the first and revised edition. *Early Childhood Research Quarterly*, 18, 427–445.
- Scarr, S., Eisenberg, M., & Deater-Deckard, K. (1994). Measurement of quality in child care centers. *Early Childhood Research Quarterly*, 9 (2), 131-151.
- Scarr, S. & Weinberg, R. (1986). The early childhood enterprise: Care and education of the young. *American Psychologist*, *41*, 1140-1146.
- Schatschneider, C., Carlson, C., Francis, D., Foorman, B., & Fletcher, J. (2002). Relationship of rapid automized naming and phonological awareness in early reading development: Implications for the double-deficit hypothesis, *Journal of Learning Disabilities*, 35, 245-256.
- Smith, M., Dickinson, D., Sangeorge, A., & Anastasopoulos, L. (2003). *Early language and literacy environment classroom observation* (Research ed.). Baltimore, MD: Brookes.
- Stipek, D., Daniels, D. Galuzzo, D., & Milburn, S. (1992). Characterizing early childhood education programs for poor and middle-class children. *Early Childhood Research Quarterly*, 7, 1-19.
- 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.
- Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., et al. (2006). Capturing quality in early childhood through environmental rating scales. *Early Childhood Research Quarterly*, 21 (1), 76-92.
- U.S. Department of Health and Human Services, Administration for Children and Families (January 2010). *Head Start Impact Study. Final Report.* Washington, DC.
- Votruba-Drzal, E., Coley, R., Maldonado-Carreño, C., Li-Grining, C., & Chase-Lansdale, P. (2010). Child care and the development of behavior problems in economically disadvantaged children in middle childhood. *Child Development*, 81, 1460-1474.
- Votruba-Drzal, E., Coley, R., & Chase-Lansdale, L. (2004). Child care and low-income children's development: Direct and moderated effects. *Child Development*, 75, 296-312.
- Watamura, S. E., Kryzer, E. M., & Robertson, S.S. (2009). Cortisol patterns at home and child care: Afternoon differences and evening recovery in children attending very high quality full-day center-based child care. *Journal of Applied Developmental Psychology*, 30 (4), 475-485.