

THE “QUALITY” OF EARLY CARE AND EDUCATION SETTINGS

Definitional and Measurement Issues

JEAN I. LAYZER

BARBARA D. GOODSON

Abt Associates, Inc., Cambridge, Massachusetts

There is a widespread belief that high-quality early care and education can improve children's school readiness. However, debate continues about the essential elements of a high-quality experience, about whether quality means the same things across different types of care settings, about how to measure quality, and about the level of quality that might make a meaningful difference in outcomes for children. Are the aspects of the child care environment that researchers measure the ones that are most strongly related to children's development? This article argues that the ways in which researchers currently measure early care environments are flawed and that the conclusions drawn about the relationship between these measures and outcomes for children are frequently incorrect or overstated. The article addresses four questions: How is the quality of the child care environment commonly defined and measured? Do the most commonly used measures capture the child's experience? Do they work well across all settings? Are researchers drawing the correct conclusions from studies that relate the child care environment to child outcomes? Finally, the article discusses some possible directions for future research.

Keywords: *child care quality; independent variable; measurement issues*

For several decades, the quality of the child care environment has been an issue of concern to policy makers, researchers, child care providers, and parents. There are three perspectives on quality: that of the parent, that of the caregiver, and that of the child. The perspectives are not mutually exclusive, but they are not the same. Parents may focus on the safety of the environment, their trust in the child care provider, and their sense that an environment meets the needs of their child. They may also take into account the degree to which the arrangement supports their employment needs (e.g., hours of operation, options for full-day or off-hours care, and gaps in operation). As children enter the preschool years, some

parents begin to see the child care arrangement as one that will help prepare the child for school.

For caregivers and teachers, however, a high-quality child care environment may have more to do with the conditions of employment. They may take into account such factors as adequate wages and benefits, including vacations and release time to acquire new skills or hone existing ones; a pleasant work environment; and reasonable staff-child ratios.

For children, high-quality child care environments offer safety and security and nurture their healthy development. In recent times, and especially with respect to children from low-income families, the developmental perspective (i.e., the child's perspective) has dominated the discussion and the research. Debates about what constitutes a good child care environment are greatly influenced by a body of research that has suggested the ability of high-quality preschool interventions to enhance the school readiness of low-income children. Much of this evidence comes from a few small-scale, carefully controlled experimental studies of educational interventions. In addition, more than two decades of correlational studies have linked the "quality" of child care (variably defined) to improved child outcomes (see, e.g., Helburn et al. 1995; Lamb 1997; Love, Ryer, and Faddis 1992; McCartney, Scarr, and Grajek 1985; National Institute of Child Health and Human Development [NICHD] Early Child Care Research Network 1999; Phillips, Scarr, and McCartney 1987; Ruopp et al. 1979).¹

Because an increasing number of young, low-income children, especially preschool-age children, are in center-based programs of some sort (either a child care center or a publicly funded preschool program), these findings have generated pressure on all forms of child care to demonstrate the capacity to enhance children's development in ways that ready them for school. Federal funding, in the form of a 4% set-aside, has flowed to states to enhance the quality of child care. Recently, a new focus on the importance of the preschool years for language development and the establishment of literacy skills has led to increased federal interest, embodied in the President's Start Smart, Grow Smart initiative, in how child care can help improve children's school readiness.

Debate continues about the essential elements of a high-quality child care environment; about whether parental definitions of quality mesh with "expert" definitions; about whether quality means the same thing across different care settings; and about the levels at which care might be damaging or, conversely, meaningfully advantageous. The new focus on the importance of the preschool years for language development and the establishment of literacy skills has led to additional examination of how a high-quality child care environment should be defined and measured.

Are the aspects of the child care environment that researchers measure the ones that are most strongly related to children's development? Do the measures that they use adequately capture those aspects? This article argues that the ways in which researchers currently measure early care environments are flawed and that the conclusions from existing studies of the relationship between these measures and outcomes for children are often incorrect or overstated. The article addresses four questions:

1. How is the quality of the child care environment commonly defined and measured?
2. Do the most commonly used measures capture the child's experience?
3. Do measures work well across all care settings?
4. Are researchers drawing the correct conclusions from studies of child care environments and child outcomes?

Finally, the article discusses some possible directions for future research.

DEFINING THE QUALITY OF THE CHILD CARE ENVIRONMENT

Quality in early care and education settings can be broadly defined as the aspects of the environment and children's experiences that nurture child development. A bewildering number of variables can be incorporated in this broad definition, and researchers disagree about their role and importance. Phillips, Scarr, and McCartney (1987) argued that staff-child ratios, staff training, and parent participation in a program are aspects of quality. Love, Ryer, and Faddis (1992) included structural aspects of the care setting, such as group size and staff-child ratio, in their definition of quality, but they viewed staff qualifications and parent involvement as contextual factors that may also influence quality.

At the core of the definition are the experiences that promote children's physical, social, emotional, and intellectual development (Layzer, Goodson, and Moss 1993). Although many aspects of the care environment directly and indirectly influence those experiences, they are not themselves aspects of quality, although they are often written and spoken of as if they were. For example, a low child-staff ratio (i.e., fewer children per caregiver) is often referred to as though, by itself, it constitutes quality. It is, however, only a possible predictor of a high-quality experience. Judged by that standard alone, the smallest informal family child care arrangement would offer the

highest quality care, because such care settings generally have fewer than three children with a single provider.

Structural characteristics. Because the structural aspects of the care environment are those that are most frequently investigated and measured, no discussion of environmental measurement can ignore them. This section briefly reviews what research tells us about the relationship of child care structure to what happens in care. The aspects of early care and education programs that research has demonstrated or experts have identified as actual or potential influences on the child's experience can be grouped into three categories: program characteristics, classroom or home characteristics, and caregiver(s) characteristics.

Program characteristics that are potential influences on the child's experience include the program's philosophy or goals and the curriculum that embodies them, parent involvement, ancillary services, and the content and intensity of staff and caregiver in-service training programs. There exists widespread belief in the importance of each program characteristic's contribution. Frequently, that belief is translated into requirements that individual programs must meet as well as funding to help them do so. For example, a substantial proportion of the funds set aside in the Child Care and Development Fund for quality enhancement is used for training and professional development of center staff and family child care providers (both regulated and unregulated). It is surprising, therefore, to realize that little or no research has assessed the importance of training and professional development as predictors of a high-quality child care experience. What research does exist is often inconsistent or contradictory. Nor, with a few exceptions (e.g., Ferrar, Harms, and Cryer 1996), have researchers attempted to translate these concepts to fit the family child care environments in which many low-income children receive care.

Classroom or home characteristics bring the observer closer to children's actual experiences. These characteristics fall into two major categories: the human composition of the classroom or home and the physical characteristics of the environment. Five elements of the composition of the classroom or home affect the child's experience and may affect the quality of that experience: the ratio of children to adult caregivers, the size of the group; the number and type of caregivers, the stability (i.e., continuity) of caregivers, and the age mixture of children in the group. The early childhood literature is replete with research studies that focus on one or more of these elements. Child-staff ratio, alone or in interaction with other elements such as group size, has frequently predicted higher quality experiences for children (e.g., an increased number of individualized interactions between

children and adults; Howes and Hamilton 1993; Layzer, Goodson, and Moss 1993; Ruopp et al. 1979; Vandell and Powers 1983; Whitebook, Howes, and Phillips 1989).

In one of the few experimental studies in this field, however, Love, Ryer, and Faddis (1992) found that a moderate negative change in child-staff ratio, which has substantial cost implications, had no significant effect on caregiver or child behavior. The National Day Care Study's (Ruopp et al. 1979) finding that group size was the most important predictor of both caregiver behavior and children's performance on standardized tests has been generally supported by later research (Phillips and Howes 1987), but the National Child Care Staffing Study (Whitebook, Howes, and Phillips 1989) failed to find a relationship between group size and global measures of the quality of the classroom environment.

Note that those two variables, like many others, may have a different meaning in home-based care, although the research findings in that area are also inconsistent. Galinsky et al. (1994) found that home caregivers who were caring for larger groups of children (three to six children) provided higher quality care than those who cared for one or two children, although contrary findings have been reported in several other studies (see Kontos et al. 1995; NICHD 1995). One reason for the inconsistent findings may be that in the case of family child care, the homes with the fewest children may also be the most informal settings, that is, relative or neighbor care by relatively untrained providers. Little research has given attention to the effects of age mixing of children, something that is the exception in center-based arrangements but is often seen in family or relative child care.

The importance of stable caregiving relationships has been stressed repeatedly in the literature on child care and early childhood (e.g., Howes and Stewart 1987). Child care advocates and policy makers have focused on the high rate of staff turnover in many centers as a threat to children's well-being. Clearly, it is difficult to build a program and integrate staff if turnover is constant. From the child's perspective, however, if the same teacher is present during the year the child spends in a group, the child has experienced continuity of care, even if the teacher leaves after the school year. Much less attention has been given to the stability of the child's presence in the care setting and the implications of instability. A recent study of the duration of child care subsidies suggests that care arrangements for some of the poorest children may be quite unstable for reasons that are not well understood. In the five states in the study, the average length of subsidy receipt was between 3 and 7 months. Loss of subsidy may have few implications for children being cared for by relatives, but for children in

more formal care arrangements, it probably results in a move to a less expensive arrangement.

Several aspects of the physical environment have been cited in the early childhood literature and appear in state regulatory codes as indicators of a high-quality environment. In addition to health and safety requirements, regulatory codes often specify age-appropriate educational materials, play equipment, and space requirements. The research literature provides partial support for the use of environmental indicators, particularly the amount and arrangement of indoor space. In general, however, specific requirements for the number and types of age-appropriate toys, materials, and equipment are based on expert opinion about what fosters the development of specific skills, rather than on research evidence.

Caregiver characteristics are perhaps the most important influence shaping a child's day-to-day experience. The qualifications of early childhood staff—experience, training, and formal education—are believed to increase the probability of competence and thus to affect the quality of the child's experience, thereby enhancing development. In center-based research, an association between formal education (particularly specialized training in early childhood education) and positive teacher behaviors has been demonstrated. The relationships are not so clear in home-based care.

Children's experiences. Children's experiences in the care setting can be divided into three categories: children's activities and groupings (the size and composition of children's groupings as well as the type of activity); the behavior of caregivers or teachers and their interactions with children and other adults; and the behavior of individual children with adults, with each other, and in solitary play.

Activities and groups answer the most basic question, "What is the child's day like?" The kinds of activities that dominate his or her experience; the size and composition of the groups of children; whether teachers or caregivers are present as monitors or active participants; and whether events occur that signal stress for children, such as conflict between children or adults or children crying for other reasons. Children should engage in a variety of activities: sometimes alone, sometimes in a small group, and sometimes with the active participation of an adult. Although there is not strong research evidence to support these beliefs, it is generally agreed that (a) children should engage in a variety of activities, (b) many of the activities for children should include the active participation and guidance of the caregiver/teacher, and (c) small group and individual activities with the caregiver provide children with the greatest opportunity to receive

high-quality adult input, especially verbal interactions. Children should not spend large amounts of time watching television; wandering aimlessly; or, at the other extreme, spending most of their time in a single, large group engaged in activities directed by the adult caregiver.

Activities and groupings provide the framework for and shape the interactions that occur between adults and children and among the children themselves. A teacher or caregiver helping a small group of children work with blocks in ways that combine fine motor activity and make-believe play has opportunities for interactions that are qualitatively different from those available to the teacher who is directing cut-and-paste activities for a class of 20 children. The interactions between the children in the small group may also be quite different from those in the larger group.

Teacher-caregiver behavior and interactions, whether in a center or a home, are critical determinants of the children's experiences. How they interact with children and how they structure their activities, their emotional tone, and the content of their interactions with the children are what primarily define a child's daily experience in the care setting. The caregivers' interest in and affection for the children, their responsiveness to children's needs, their use of positive guidance and disciplinary techniques, and their care in monitoring activities all contribute to creating a safe and nurturing environment. These are all aspects of caregiver behavior that developmental experts and parents would agree are essential.

Caregivers can and often do act as teachers, but how that role is defined may differ according to the ages of the children and the setting. Caregivers can help preschool-age children learn social skills they will need later, such as taking turns, sharing, and being empathetic. They can support children's fine-motor and active physical play, both of which are crucial avenues for development. Finally, they can support children's acquisition of knowledge and intellectual skills. Although the latter may take place more consciously in a center setting, children's daily lives provide constant opportunities for learning, through the child's own experimentation and play alone as well as with other children. Adult caregivers can ensure that opportunities for such learning occur, whether in a home or a center.

Children's behavior and interactions are the final element. A description of the physical characteristics and activities that go on in the classroom or home, coupled with a description of caregiver or teacher behavior, can provide a sense of what life is like, *on average*, for children in that setting. But it cannot tell us much about the experience of an individual child. For that, one must understand what that child does in care and the type and amount of his or her interactions with other children and with adults. As we showed

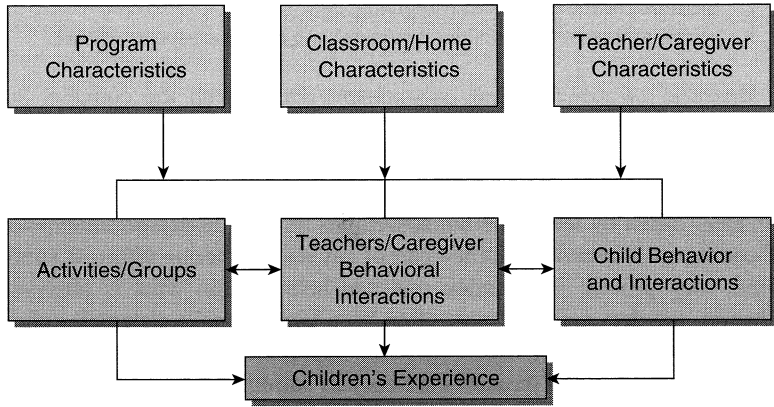


Figure 1: Influences on the Child's Experience in Early Child Care and Education

in an earlier study (Layzer, Goodson, and Moss 1993), those experiences may differ greatly among children: Some may get a good deal of attention from the caregiver, whereas others receive very little.

Figure 1 is a diagram of the pathways through which aspects of the care environment influence the child's experience in care. The kinds of activities in which children are engaged; the size and composition of groups; the types and nature of caregiver interactions with children; and children's interactions with adults, peers, and materials in the environment are all shaped to a greater or lesser extent by aspects of the program, classroom or home, and caregiver. The importance of any aspect may differ, depending on whether care is in a center or in a home. In turn, the activities and groupings may influence both the ways in which caregivers interact with children and the ways in which individual children behave with adults, peers, and materials in the environment. Taken together, the mixture of activities and interactions make up the child's experience in care.

MEASURING PROGRAM CHARACTERISTICS AND PROCESSES

How, then, does one measure the aspects of the care setting that influence or constitute children's experience? The measures that are currently used fall into three general categories:

1. measures of structural and environmental characteristics of the care,
2. measures of the process of caregiving, and
3. global measures that combine structural and dynamic aspects of care.

Measures of structure and environment. Faced with the complex and expensive task of measuring children's experience, many researchers have adopted the strategy of gathering information about centers, classrooms, and homes through survey methods, such as interviews with parents and providers (often by telephone but sometimes through in-person interviews). Of course, such interviews form a part of any data-gathering effort; at a minimum, they provide information that can help explain what is observed. However, the data gathered have often been used to make statements about quality, especially in large-scale research. The reasons for this approach are not hard to discern; the measures are easy to administer and relatively inexpensive to use. Although there is, increasingly, an understanding that the measures are, if anything, mediators of the care process rather than direct indicators, a few of them are consistently related to other, more direct measures of what actually happens in the child care setting. Other drawbacks to the use of interviews are that their relationship to other measures may differ, depending on the type of setting, and that no structural measures have been identified that consistently predict the quality of care in home-based settings.

Measures of caregiving process. At the other end of the spectrum are measures that attempt to capture the dynamics of care: the activities in which children and caregivers are engaged, the nature of the caregiver's interaction with children, and the nature of children's interaction with adults and other children as well as their behavior in solitary play. As stated earlier, these dynamic processes, taken together, represent the quality of the early childhood experience, according to most researchers. To measure the dynamics accurately would be to capture the essence of the child's experience. These measures, however, tend to be more narrowly focused, assessing one aspect of the process in depth. In reality, measures of classroom process are hardly ever used in large-scale research. Again, the reasons are not difficult to understand. It is not possible to capture all aspects of a child's experience in a single instrument, which is what most researchers would prefer, given the cost and difficulty of training observers to use multiple instruments. Even a measure designed to assess one aspect of the process, such as teacher behavior, is difficult and expensive to train and administer. The most expensive and difficult measures are those that attempt to capture the experience of individual children.

Global quality ratings. Given the expense and difficulty of using measures that focus squarely on classroom process, along with the concern that just measuring structural attributes of the environment is inadequate, it was probably inevitable that the attention of researchers would turn increasingly to measures that combine structural and process characteristics, often in a single item on a rating scale. The measures fall between the two extremes in terms of the ease and expense of training research staff and administering them, and they are easier and less expensive to analyze than the two types of measures described earlier.

The limitations of all three types of measures, including the most sophisticated process measures, are (a) they are usually one-time snapshots of the care setting; (b) they do not capture the experience of the individual child; and (c) they are easier to develop for more formal settings, so that assessment of informal settings has lagged behind.

A one-time snapshot is the result when the instrument measures what happens on a particular day. The activities and interactions in any setting may vary greatly over a week or a month, but our research procedures for the most part ignore this basic fact. None of the types of measures described above capture the stability of the experience over time, that is, how long the child is in the care setting or is with the same caregiver. Little is known about the independent contribution that stability of care makes to children's development, but the value to the child of a high-quality child care experience is almost certainly affected by the length of time the child is exposed to the experience.

The individual child's perspective is not necessarily captured when the instrument measures what is happening to the classroom as a whole. One of the greatest weaknesses of most observation measures is that they do not necessarily capture the experience of individual children. The average score for a classroom or center may be a useful prompt for self-improvement, but it is a blunt measure to attach to the achievement score or socioemotional rating for an individual child. It is as though one used the average educational level of the mothers of all children in a center as a covariate for an individual achievement score. This weakness is important because, as anyone who has observed a classroom knows, and as our previous research has shown, individual children can have very different experiences in the same classroom.

For example, in an observational study of early childhood programs (Layzer, Goodson, and Moss 1993), staff spent two thirds of their time in direct interaction with children. When the focus of the observation shifted to individual children, however, in almost half of the classrooms, some children received no individual attention over the two mornings of observation. Across

all the classrooms, almost one third of the children received no individual attention from an adult during the observations.

Individual children also have different experiences playing or working with materials alone or in interaction with other children. Global measures are not able to detect these important differences in children's experiences.

Assessing informal settings. The public dialogue about the importance of high-quality child care frequently focuses on center-based programs, as though that is where most children receive care. Only about half of all child care is in centers (West, Wright, and Hausken 1996; Capizzano, Adams, and Sonenstein 2000). For a variety of reasons, low-income families are more likely than other families to rely on informal care arrangements with relatives and friends (Capizzano, Adams, and Sonenstein 2000; Casper 1997). In addition to its greater affordability, care by relatives and friends may be the only kind of care that meets the needs of parents who work off hours or irregular schedules or whose employment may be unstable.

Many families who use kith-and-kin care do so because they prefer to rely on relatives and friends whom they know and trust. When they are asked why they chose a relative, they will often cite common values and beliefs, the importance of an arrangement that looks like home, the ability to have siblings in the same setting, and the support that friends and relatives offer to them as well as their children. None of the most commonly used measures of the child care environment capture any of the aspects of informal settings that parents cite as most important. Little effort has been made to think through what should be included in care provided by neighbors during the evening or by grandmothers after school.

WHAT ARE THE LIMITATIONS OF GLOBAL MEASURES?

Recognizing that all types of measures share the drawbacks described above, we believe that global measures of quality, increasingly used for both research and program improvement, present an additional set of problems. We have chosen to focus our discussion of these issues on the Early Childhood Environment Rating Scale—Revised Edition (ECERS-R; Harms, Clifford, and Cryer, 1998) and its companion measures, the Infant-Toddler Environment Rating Scale—Revised Edition (ITERS-R; Harms, Cryer, and Clifford, 2003) and the Family Day Care Rating Scale (FDCRS). (Our comments apply to the original ECERS and ITERS as well as their revised editions.) Taken together, this family of measures is the one most widely used and has come to dominate the discussion of child care quality.

In addition to their wide use in research studies, they are increasingly used by state and local child care agency staff to assess the quality of child care settings. ECERS scores have been used as evidence that child care is, on the whole, "mediocre" (not a term used in the scale itself) as the basis for differential payments to providers; and as justification for investing additional money in quality improvement, among other things. What do these instruments measure, and how do they do it?

The ECERS-R (and its companion measure, the ITERS-R) is a 43-item scale that allows the observer to rate six aspects of the environment. Each item in the scale blends structural aspects of the environment and behavioral indicators. For example, an item might ask about specific materials in the room, their accessibility, and the caregiver's use of the item. Each item is scored on a 7-point scale from *inadequate* to *excellent*, and a total score is computed to give an environmental rating.

On the ECERS-R, a classroom would receive a score of 3.0 (*minimal*) if it had some appropriate equipment for building gross motor skills that was seldom in use; it would receive a score of 5.0 (*good*) if the gross motor equipment were readily available and sturdy, stimulating a variety of skills, and if building and dramatic play equipment were included in gross motor areas. If one or part of one of these statements could not be made (e.g., if dramatic play equipment were not included in the gross motor area), then the classroom would receive a score of 4.0. On an item assessing opportunities for teachers' professional growth, a center (this is not a classroom-level item) would receive a 3.0 rating if it had a limited professional library, staff meetings were limited to administrative issues, and no in-service training were provided; it would receive a 5.0 if it had a good professional library that had current materials on a wide variety of subjects readily available, regular staff meetings that included staff development activities, and a plan for orienting new staff. A center with a limited professional library but all the other components of the 5.0 rating would score 4.0 on this item.

Figure 2 shows the areas covered by the subscales of the ECERS-R and the number of items that contribute to each subscale. The seven subscales carry the same weight in the total score.

The scale was originally developed as a tool that centers could use for self-assessment to target areas for improvement. As Figure 2 shows, its scope is broad, covering aspects of the environment important for the comfort of staff and parents as well as children. The scale is based primarily on expert opinion and reflects a generous and expansive vision of what is necessary to create a comfortable and nurturing center environment for children.

Given its many positive aspects, is there any reason to believe that the ECERS-R and other global measures are not the ideal way to assess

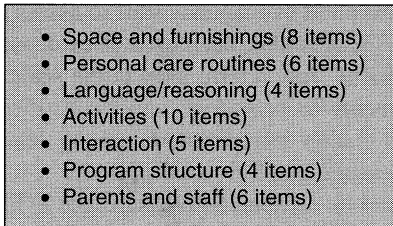
-
- 
- Space and furnishings (8 items)
 - Personal care routines (6 items)
 - Language/reasoning (4 items)
 - Activities (10 items)
 - Interaction (5 items)
 - Program structure (4 items)
 - Parents and staff (6 items)

Figure 2: Subscales of the Early Childhood Environment Rating Scale-Revised Edition

children's experience? Several important issues involve what the instrument does (and does not) measure, how it is used (or misused), and the limits on what can be learned from such a measure:

- the underlying theory about child development that informs the content of the measure;
- the mixture of items measured and how observers make their judgments;
- the concealment or absence of critical aspects of children's experience;
- the psychometric qualities of the ECERS-R and their relevance to the use of the ECERS-R rating in measuring changes in the quality of early childhood programs; and
- the ways in which global rating scores are linked to outcomes.

The sections that follow briefly discuss each of those issues.

Underlying theory. A high quality rating on the ECERS-R depends to a large extent on two aspects of the environment: (a) the richness and accessibility to children of the materials in the classroom and (b) the extent to which the program meets standards for children's health and hygiene practices. In terms of children's use of materials in the classroom, the ECERS-R is built on the hypothesis that the highest-quality learning experiences for children are those that involve independent play with child-selected materials. In fact, on the ECERS-R, involvement of the teacher in selecting materials (e.g., selecting which puzzles are put out on tables, selecting which art materials are brought outside for use in outdoor children's play) counts against the program's quality rating. This perspective on developmentally appropriate practice is very consistent with the 1985 version of

the National Association for the Education of Young Children (NAEYC) guidelines for quality in early childhood education. However, the field has experienced a significant shift in perspective since the 1980s. Research on the impacts of early childhood interventions has shown that even the highest quality early childhood settings, as defined by their ECERS-R scores, have not been successful in eliminating the gap in school readiness between low-income or at-risk preschoolers and their better off peers.

The newest version of the NAEYC guidelines recommends a much more active role for the teachers, especially in providing additional language stimulation and instruction in early literacy knowledge. The revisions to the original measure reflected in the ECERS-R place slightly more emphasis on language and literacy specifically. These revisions, however, may not meet the field's standard for assessing critical aspects of the classroom experience for children from backgrounds that are not rich in adult-child language instruction or print exposure. One researcher in the field has suggested that the ECERS, despite its advantages, provides "woefully slight information about aspects of classroom that are directly linked to supporting early literacy" (Dickinson, 2002, 28).

The second strong theme in the ECERS is its focus on health and hygiene practices. One rationale is that programs with poor adherence to hand-washing and disinfecting standards may have higher rates of child illness, which in turn leads to higher rates of absences. Children who are frequently sick and consequently absent are less likely to receive the full benefit of a high-quality early childhood environment.

A second rationale is the contention that a program's practices in the area of health and hygiene are significantly correlated with harder to measure dynamic aspects of quality, such as the content and quality of the verbal interactions in the classroom. This means that measuring health and hygiene practices is a valid way to get at dynamic quality. The research evidence for this hypothesized correlation between health practices and classroom quality is not clear. Many in the field are disconcerted by the proportion of ECERS items that require close examination of the details of a program's practices, such as whether the liquid used to wash tables contains disinfectant, the toileting behaviors of individual children, and the like. In many respects, absent strong evidence of a link between hygiene and the dynamic characteristics of the classroom interactions, the items on hygiene would seem more appropriately the province of health inspectors rather than observers of the quality of the educational environment for children.

Mixing resources and interactions. Each item on the ECERS-R scale bundles material resources and how they are used by the teacher or the

children. Increasingly, researchers are concerned that the measure overvalues some resources and aspects of the physical environment and captures poorly, if at all, two aspects of teacher behavior linked to children's development: responsiveness to children and use of language. Some researchers report evidence of a "halo" effect: Classrooms rated highly on one subscale tend to receive high scores on the others. It is as though the structure of the measure leads raters to generate a single, unified view of the classroom. In studies that use both global measures and measures that assess interactions among teachers and children, classrooms are consistently rated higher on the global measures than on the interaction measures, an indication that the raters focus more on the physical aspects of the environment than on the critical interactions that take place there.

Assessing high-quality language and literacy practices. In addition to their inability to capture the experiences of individual children or to capture unique aspects of informal care settings, the most commonly used global measures can conceal critically important areas of program weakness. The past decade has seen an increasing emphasis on the importance of early language and literacy development and on the role of early care and education programs in promoting them. The NAEYC has reversed its earlier position on direct literacy instruction in response to three decades of research that provides evidence about the importance of early support of children's language growth, engagement with print materials, and literacy-related activities. Current measures offer limited coverage of these critical aspects of the child's experience, and they do not offer a means to weight their importance. For example, the four ECERS items that measure the extent to which the environment meets adult needs make the same contribution to the total score as the four items that assess the quality of children's language and reasoning experiences. As a result, it is possible to give a highly favorable rating to programs that offer minimal support for language and literacy acquisition.

Psychometric properties. There are at least three concerns about the psychometric properties of the ECERS-R. First, the points on the scale of measures like the ECERS-R may not be equidistant from each other. The difficulties that centers experience in moving beyond a score of 5 (good) to 7 (high quality) suggest that the step from 5 to 7 is much larger than the step from 3 (mediocre) to 5.

Second, one gap in our knowledge about the psychometrics of the ECERS-R and its companion measures (and other ordinal rating scales) involves the properties of the scale itself. The administration of the measure

mandates stop points where, if a program fails at a low point on the scale, higher-level items are not evaluated. The implicit hypothesis is that no program that gets a low score on an item would ever meet any of the higher-level criteria. There are no data to support the setting of these stop points (i.e., data that would show that it is rare for a program to fail on a lower component of an item yet still demonstrate some of the higher-quality practices included in that item).

Another part of the psychometric information that is needed is data on the statistical coherence or validity of the subscales (i.e., the extent to which the items hang together and contribute independently to the overall subscale score). Current statistical methods, such as Item Response Theory (ITR) scaling, would test the reliability of the subscales.

Finally, the ECERS-R is frequently used to measure change in quality over time, for instance, when there are initiatives in programs aimed at improving their quality. In many of these initiatives, change on the ECERS-R of a point is the criterion for significant improvement. However, in establishing rater reliability in the field on the ECERS-R, raters are considered to be reliable if their scoring of a program is within 1 point (on a scale of 1 to 7) of the score given the same program by a trained (criterion) coder. This means that the change of a point on the ECERS-R may confound unreliability and any true improvement, which undercuts the validity of the results in assessing change over time.

Linking global ratings to child outcomes. Spending by state and local governments on initiatives to improve the quality of child care has been greatly influenced by studies that have linked scores on global ratings to outcomes for children. The most visible and influential of these studies was the Cost, Quality and Child Outcomes (CQO) Study, conducted by a consortium of respected academic researchers in the mid-1990s. A major finding of the study was that "children's cognitive and social development are positively related to the quality of their child care experience" (Helburn et al. 1995, 26), a finding that echoed those of a number of other studies. Another widely cited finding was that good quality costs more, but not much more. The study estimated that an expenditure of an additional \$0.16 per hour would be needed to raise the quality of the child care environment from *medium* (4.0) to *good* (5.0), and it strongly recommended that that be done.² The problem is that, to find even the small but significant association between scores on the global ratings of the child care environment and child performance on standardized tests, the researchers compared children in centers rated as being of the lowest and highest quality, respectively, presumably because they found no

significant differences between children who attended “medium-quality” versus “good-quality” centers although these latter points are those on which the cost-quality analysis and policy recommendations focused.

The foregoing is not to suggest that the quality of the child’s experience in an early care and education setting is unimportant; but because the global measures are so broadly focused, are not strongly linked to the outcomes of interest, and do not reflect the experience of individual children, they may actually understate the relationship between the quality of the experience and outcomes. As things stand, researchers have too often overstated the implications of small differences in environmental rating for children’s development.

Policies based on the findings from the CQO Study and other research that relied heavily on global measures of the environment are not likely to yield the hoped-for results. Such policies could actually have the negative effect of further constraining parental child care options. Most activities aimed at increasing child care quality have the concomitant effect of raising the cost of that care. When parents pay for care, the result may be to place some types of child care beyond their reach. When public funds pay for child care, fewer children can be served at a fixed level of resources. If and when a compelling case can be made that (a) care of a specific quality has a meaningful (i.e., substantial and lasting) impact on the school readiness and achievement of children who are otherwise at risk for academic failure and that (b) effective and reliable strategies can be identified to raise the quality of existing care to the level required to achieve such effects, a strong argument could be mounted for greatly increased public investment.

OTHER WAYS TO ASSESS THE ENVIRONMENT

If the global ratings most commonly used are not always appropriate, do other measures offer a better look at variables that matter? Several instruments deserve to be considered as alternatives to the ECERS-R and, less frequently, to the ITERS-R. Some possible candidates include the following:

- The Observational Rating of the Care Environment (ORCE) Qualitative Ratings (NICHD Early Child Care Research Network 1995) measures caregiver behavior including sensitivity and responsiveness to nondistressed communication, detachment, intrusiveness, stimulation of cognitive development, positive regard for the child, negative regard for the child, flat affect, and sensitivity to distress. This measure captures well the affective and responsive

aspect of caregiver behavior that is an essential component of good caregiving. (Note that this is not the child-focused ORCE measure, which is a complex micro-observation.)

- The Teacher Involvement Scale (Howes and Stewart 1987) is a time-sample measure of the specific kinds of interactions that occur between a caregiver and a child, from ignoring to simple contact to intense contact.
- The Supports for Early Literacy Assessment (SELA) is a new tool still under development by Sheila Smith at New York University. The SELA assesses the quality of support for young children's literacy development in center-based preschool settings; it combines a classroom observation and an interview with the caregiver.
- The Preschool Classroom Implementation Rating Scale (Frede, 2002) is a rating scale that assesses the extent to which preschool staff extend children's activities and problem solving and children's thinking.
- The QUEST measurement battery (Goodson, Layzer, and Layzer, 2005) is a battery of measures that assess separately the behaviors and interactions of caregivers with children, the richness of materials and resources in the environment, children's activities and groupings, and the interactions of individual children with each other and with the caregiver. The measures are broad in scope but emphasize aspects of care that are related to positive child outcomes. The measures were developed for use in the National Study of Child Care for Low-Income Children (Layzer and Goodson final report in preparation), can be used as a group or as individual measures, and are appropriate for use across all types of setting. For the study, they were used in combination with a measure of caregiver affect and responsiveness.
- The Observation Measures of Language and Literacy Instruction (the OMLIT; Goodson, Layzer, Smith, and Rindzius, 2005) is a battery of measures that focus on the instructional practices and literacy resource in early childhood education classrooms. The battery was developed for the fourth national evaluation of the Even Start Program (CLIO: the Even Start Classroom Literacy Interventions and Outcomes Study), which is funded by the U.S. Department of Education.³ The battery includes five measures: a snapshot of the literacy activities and oral language in the classroom every 15 minutes; a detailed description of the content and instructional practices of staff in any literacy activities, also recorded every 15 minutes; a detailed description of the adult's strategies in reading aloud to children, especially the quality of the discussion with children before, during, and after the book is read; a description of the literacy resources in the classroom; and an overall rating of the quality of the classroom practices in promoting children's oral language, phonological knowledge, and emergent writing and practices with bilingual children.

Similar measures are under development to assess the extent to which the care environment supports young children's understanding of mathematics. No single measure is sufficient by itself to represent all that should

be captured in an assessment of quality, but in combination, the measures focus more sharply than the global measures on some aspects of care for young children that can be directly linked to their readiness for school.

WHAT NEXT?

Publicly funded child care is under increasing pressure to address the bipartisan goal of school readiness and, more specifically, to lay the language and literacy foundation for future school success. Little discussion has taken place about whether this mission is feasible for all types of child care or whether other strategies need to be adopted to meet the needs of children who spend all or part of their day in informal care arrangements. This discussion would be assisted by the development of well-specified models of the pathway to school readiness that identify critical elements of the early care and education environment. Those elements could then be built into observational measures to assess the environment. Some efforts of this kind are already under way. In addition, if evaluations of care settings are going to continue to be linked to the performance of individual children, then researchers must find ways to incorporate children's perspectives and experiences into measures of quality.

NOTES

1. The issue of selection bias, which plagues all of these studies, continues to be a focus of debate in the research community (see, e.g., Vandell and Wolfe 2000).

2. Given the problems of selection bias, very low cumulative response rates, and attrition from the sample that the study faced, its findings (and recommendations) should have been viewed with more caution than was the case.

3. For information about the Observation Measures of Language and Literacy Instruction (OMLIT) measures, contact Tracy Rimdzius, U.S. Department of Education, Tracy.Rimdzius@ed.gov.

REFERENCES

- Capizzano, J., G. Adams, and F. Sonenstein. 2000. *Child care arrangements for children under five: Variation across states*. Assessing the New Federalism: National Survey of America's Families, Series B, No. B-7. Washington, DC: Urban Institute.

- Casper, L. M. 1997. *Who's minding our preschoolers?* Fall 1994 Update, Current Population Reports, Household Economic Studies, P70-62. Washington, DC: U.S. Department of Commerce.
- Dickinson, D. K. 2002. Shifting images of developmentally appropriate practice as seen through different lenses. *Educational Researcher* 31:26-32.
- Ferrar, H. M., T. Harms, and D. Cryer. 1996. *Places for growing: How to improve your family child care home*. Princeton, NJ: Mathematica Policy Research.
- Frede, E. (1989). *The Preschool Classroom Implementation Rating Scale*. Newark, NJ: Rutgers University.
- Galinsky, E., C. Howes, S. Kontos, and M. Shinn. 1994. *The study of children in family child care and relative care: Highlights of findings*. New York: Families and Work Institute.
- Goodson, B. D., J. Layzer, and C. Layzer. 2005. *QUEST observation system: Environment rating and provider rating*. Cambridge, MA: Abt Associates, Inc.
- Harms, T., R. M. Clifford, and D. Cryer. 1998. *Early Childhood Environment Rating Scale—Revised Edition*. Carrboro, NC: Frank Porter Graham Child Development Institute.
- Harms, T., D. Cryer, and R. M. Clifford. 2003. *Infant/Toddler Environment Rating Scale—Revised Edition*. Carrboro, NC: Frank Porter Graham Child Development Institute.
- Helburn, S., M. L. Culkin, J. Morris, N. Mocan, C. Howes, L. Phillipsen, D. Bryant, E. Clifford, D. Cryer, E. Peisner-Feinberg, M. Burchinal, S. L. Kagan, and J. Rustici. 1995. *Cost, quality, and child outcomes in child care centers*. Denver: University of Colorado.
- Howes, C., and C. Hamilton. 1993. The changing experience of child care: Changes in teachers and in teacher-child relationships and children's social competence with peers. *Early Childhood Research Quarterly* 8:15-32.
- Howes, C., and P. Stewart. 1987. Child's play with adults, toys and peers: An examination of family and child care influences. *Developmental Psychology* 23:423-30.
- Kontos, S., C. Howes, M. Shinn, and E. Galinsky. 1995. *Quality in family child care and relative care*. New York: Teachers' College Press.
- Lamb, M. E. 1997. Nonparental child care: Context, quality, correlates, and consequences. In *Handbook of child psychology*. Vol. 4, *Child psychology in practice*, 5th ed., ed. I. Sigel and A. Renninger (vol. eds.), 73-133. New York: John Wiley.
- Layzer, J., B. D. Goodson, and M. Moss. 1993. *Life in preschool: Final report of the Observational Study of Early Childhood Programs*. Cambridge, MA: Abt Associates.
- Love, J. M., P. Ryer, and B. Faddis. 1992. *Caring environments—program quality in California's publicly funded child development programs: Report on the legislatively mandated 1990-91 staff/child ratio study*. Portsmouth, NH: RMC Research Corporation.
- McCartney, K., P. D. Scarr, and S. Grajek. 1985. Day care as intervention: Comparisons of varying quality programs. *Journal of Applied Developmental Psychology* 6:247-60.
- National Institute of Child Health and Human Development. 1995. *Child care in the 1990s: The NICHD study of early child care*. Indianapolis, IN: Society for Research in Child Development.
- National Institute of Child Health and Human Development Early Child Care Research Network. 1999. Child outcomes when child care center classes meet recommended standards for quality. *American Journal of Public Health* 89:1072-77.
- Phillips, D. A., and C. Howes. 1987. Indicators of quality in child care: Review of research. In *Quality in child care: What does research tell us?* ed. D. A. Phillips. Washington, DC: National Association for the Education of Young Children.
- Phillips, D., S. Scarr, and K. McCartney. 1987. Child care quality and children's social development. *Developmental Psychology* 23:537-43.
- Ruopp, R. R., J. Travers, F. Glantz, and C. Coelen. 1979. *Children at the center: Final report of the National Day Care Study*. Cambridge, MA: Abt Books.

- Vandell, D. L., and C. P. Powers. 1983. Day care quality and children's free play activities. *American Journal of Orthopsychiatry* 53:493-500.
- Vandell, D. L., and B. Wolfe. 2000. *Child care quality: Does it matter and does it need to be improved?* Madison: University of Wisconsin, Institute of Research on Poverty.
- West, J., D. A. Wright, and E. G. Hausken. 1996. *Statistics in brief: Child care and early education program participation of infants, toddlers, and preschoolers*. Washington, DC: National Center for Educational Statistics.
- Whitebook, M., C. Howes, and D. A. Phillips. 1989. *Who cares? Child care teachers and quality of care in America*. Final Report of the National Child Care Staffing Study. Oakland, CA: Child Care Employee Project.

Barbara Goodson is principal scientist at Abt in education and child development, a company-wide resource on design and instrumentation for evaluating outcomes of educational and social interventions for parents and children. On a number of nationally important research studies, she has developed measures of (a) the quality of instructional practices in early childhood settings, (b) parent outcomes, (c) child and youth outcomes, and (d) the quality of program services. She is directing studies of a range of early childhood language and literacy interventions in sites nationwide. She led the development of state-of-the-art instrumentation on language and literacy instructional practices in early childhood education.

Jean Layzer has worked in early care and education research for more than 30 years. She recently directed two major child care studies: the national Study of Child Care for Low-Income Families, a 7-year policy study conducted in 17 states and 25 communities, and the Evaluation of Child Care Subsidy Experiments, both sponsored by ACF. The latter study involves the design and implementation of a series of experiments to test aspects of child care subsidy policy, including the use of quality set-aside dollars. One of the experiments, conducted in partnership with the Miami-Dade School Readiness Coalition, is testing the feasibility and impacts on teachers and children of training low-paid, child care center staff to implement curricula focused on developing language and preliteracy skills.