

CENTER FOR THE STUDY OF

# Child Care Employment

## **Evaluating The Early Childhood Environment Rating Scale (ECERS): Assessing Differences Between the First and Revised Edition**

Laura Sakai, Ph.D.  
Honolulu, Hawaii

Marcy Whitebook, Ph.D.  
University of California, Berkeley

Alison Wishard and Carollee Howes, Ph.D.  
University of California at Los Angeles

Running head: EARLY CHILDHOOD ENVIRONMENT RATING SCALE

Evaluating The Early Childhood Environment Rating Scale (ECERS):

Assessing differences between the first and revised edition

Laura M. Sakai, Ph.D.

Marcy Whitebook, Ph.D.

Honolulu, Hawaii

University of California at Berkeley

Alison Wishard and Carollee Howes, Ph.D.

University of California at Los Angeles

Short title: Evaluating The Early Childhood Environment Rating Scale

Acknowledgment: This research was funded by grants from the David and Lucile Packard Foundation to the Center for the Child Care Workforce and the Center for the Study of Child Care Employment at the University of California at Berkeley. The authors wish to thank the research assistants and our Advisory Committee members and reviewers. A special thanks to the child care teachers and directors who participated in this study, welcomed us into their centers, and gave so generously of their time.

Correspondence concerning this article should be addressed to Laura Sakai, 3357 Pakanu Place, Honolulu, Hawaii 96822, Email: lsakai01@hawaii.rr.com.

### Abstract

Before 1998, most large-scale studies of center-based child care programs measured quality using the 1980 version of the ECERS. To know whether data from studies conducted after 1998 using the revised ECERS-R can be fairly compared to data from studies using the 1980 ECERS, simultaneous assessments using both measures in a sample of 68 classrooms were conducted. The results suggest that the original ECERS and ECERS-R can be viewed, as their authors intended, as comparable measures of quality. Scores were highly correlated and similarly distributed. Principal components analysis resulted in two factors for both measures. Both measures fall short in addressing staff stability and key components of culturally sensitive practice, such as communicating with families in their home language.

Evaluating the Early Childhood Environment Rating Scale:  
Assessing differences between the first and revised edition

In the last thirty years, considerable research has documented the mediocre quality of many of our nation's child care programs and their ineffectiveness in optimizing children's development (Coelen, Glantz, & Calore, 1979; Helburn, 1995; Keyserling, 1972; Phillips, Mekos, Scarr, McCartney, & Abbot-Shim, 2000; Whitebook, Howes, & Phillips, 1990). Growing concern about poor child care quality has prompted policy makers and early education professionals to develop and fund a variety of strategies to enhance child care quality (Bellm, Burton, Shukla, & Whitebook, 1997; Whitebook & Eichberg, 2002; Whitebook, Sakai, & Howes, 1997). Millions of public and private dollars have been dedicated in recent years to enhancing the quality of both center- and home-based child care (Whitebook & Eichberg, 2002; Whitebook et al., 1997). Given the importance of high-quality child care to children's developmental well-being, as well as the large public investment targeted toward improving care, the way in which researchers define and measure quality is under renewed examination (Chang, Muckelroy, & Pulido-Tobiassen, 1996; Holloway, Kagan, Fuller, Tsou, & Carroll, 2001; Lamb, 1998; Love, Schochet, & Mechstroth, 1996; National Research Council and Institute of Medicine, 2000; Phillips & Howes, 1987).

Prior to 1998, the Early Childhood Environment Rating Scale (ECERS) developed by Harms and Clifford was the most widely used assessment of global child care classroom quality in center-based programs. Most large-scale studies, such as the Cost, Quality and Child Outcomes in Child Care Centers study (CQCO) and the National Child Care Staffing Study (NCCSS) have used the ECERS to measure classroom and global center quality (Helburn, 1995; Whitebook et al., 1990).

In addition to its use as a research instrument, the Early Childhood Environment Rating Scale (ECERS), and later its revision, the ECERS-R, (Harms & Clifford, 1980; Harms, Clifford, & Cryer, 1998) increasingly have been used as a self-assessment tool by teaching and administrative staff in child care centers. For example, the California Early Childhood Mentor Program, which currently operates at more than 70 community colleges throughout the state, employs ECERS-R ratings in its selection of mentor teachers (California Early Childhood Mentor Program, 2001). In addition, many initiatives designed to improve child care compensation and professional development, such as California's Compensation and Recognition Encourage Stability (CARES) program, require an ECERS-R classroom assessment as a condition for ongoing participation by teachers (Burton, Mihaly, Kagiwada, & Whitebook, 2000). In North Carolina, centers' licensing status and reimbursement rates for serving children of low-income families are based on ECERS-R scores (North Carolina Department of Social Services, 1999).

Published in 1980, the original ECERS is an observational instrument that focuses on the day-to-day quality of classroom environments, activities and interactions (Harms & Clifford, 1980). Researchers typically spend two or more hours in a classroom conducting an ECERS assessment. The measure includes several items arranged in seven content areas, as shown in Table 1, and is used to assess whether a classroom has developmentally appropriate supplies and activities and whether teaching staff use developmentally appropriate practices when interacting with children during activities related to each area.

Widespread use of the ECERS as a research and assessment tool, along with changes in thinking about best practices for children, prompted a revision of the ECERS in 1998 (Harms et al., 1998). The authors sought to balance continuity and innovation in the revision, and examined

several types of information to guide the revision process. They conducted content analyses of the ECERS with other measures of global quality, extensively reviewed data from other studies using the ECERS, and gathered feedback from ECERS users. With respect to the latter, they conducted three focus groups with practitioners and researchers to explore how the ECERS functioned in settings with regard to inclusion of children with disabilities and cultural sensitivity (Harms et al., 1998). According to the instruments' authors, the ECERS and the revised version (ECERS-R) share the same rationale and underlying constructs, both focusing on the global definition of the classroom environment with quality determined by our current knowledge of "best practices." The revised instrument is scored using the same seven-point scale as the original. The ECERS-R differs from the original ECERS slightly in its seven subscales, however, as shown in Table 1.

Additional changes included: elimination of the infant/toddler items because of the development of the Infant/Toddler Environment Rating Scale (ITERS; Harms, Cryer, & Clifford, 1990); shifts in various aspects of the formatting and scoring to make ECERS-R consistent with other scales developed by the same authors (e.g., the ITERS, the Family Day Care Rating Scale, Harms & Clifford, 1989) or easier to use; elimination of certain redundant items; separation of some items into several to deepen content; inclusion of new areas not covered in the original, such as use of video and computers; and the addition of indicators and examples throughout the measure to make it more inclusive and culturally sensitive. Using two sets of field tests, the authors assessed the inter-rater reliability of the new measure and found quite high levels of agreement (.92 Pearson product moment and .86 Spearman rank order correlations), comparable to those of the original measure (Harms et al., 1998). Internal consistencies of the scale at the subscale and total score levels were found to have reasonable levels of internal agreement.

Interrater intra-class correlations ranged from .71 to .88 with a total scale internal consistency of .92 (Harms et al., 1998). At the time of publication of the revision, many questions about reliability and validity between the two measures remained unanswered, including the degree to which the revised version maintains the same magnitude of scores as the original version, whether the factor structure of the original scale and the revised scale are similar, and whether the two versions predict child development outcomes similarly.

While researchers and practitioners welcomed the revision of the instrument, the changes posed a dilemma for those, like us, conducting longitudinal research and/or wishing to compare current ratings of quality with previous studies. Because we were engaged in a longitudinal study which used the original ECERS as our measure of global quality, we were faced with the decision of whether to use the original version or to switch to the revised version in our third and final round of data collection. Results for the first and second rounds of data collection were published in 1997 in a report entitled, *NAEYC Accreditation as a Strategy for Improving Child Care Quality: An Assessment* (Whitebook et al., 1997). This study examined the degree to which centers seeking NAEYC accreditation improved in quality over a two-year period (1994-1996), the level of quality such centers achieved, and the types of support that helped centers in the NAEYC accreditation process. In 2000, we sought to ascertain the extent to which the centers in our sample sustained improvements in quality made in the process of seeking accreditation, and the impact of the staffing crisis on the overall quality of their services. Once again, we sought to observe centers using the ECERS as our measure of quality. Because reliability studies on the two measures were not yet available when we began data collection in 2000, we explored the implications of using one or the other measure. If we used the original version, it might be difficult to compare our results with current and future research. If we used the new version, and

it was not comparable to the original, comparisons with our previous data would be compromised. Based on discussions with our funding agency, we decided to use both measures, recognizing the increased costs of conducting two assessments. This allowed us to perform reliability tests between the original and revised ECERS, the results of which are the focus of this article.

Specifically, this study addressed two main issues. First, we sought to understand to what extent the original and revised versions of the ECERS were equivalent measures of quality. To address this question, we looked at three issues of comparison:

1. To what extent are global ratings conducted simultaneously in classrooms using the ECERS and ECERS-R equivalent? In other words, can researchers and practitioners assume that a rating of 5 on the ECERS would likely be a rating of 5 on the ECERS-R?
2. Do both the original and revised ECERS have the same empirical dimensions as measured by their factor structure? Earlier research using the ECERS has provided one factor focusing on caregiver interactions and another on the environment and activities (Helburn, 1995; Whitebook et al., 1990).
3. To what extent are correlations between the ECERS-R and measures of teacher behavior similar to those for the original ECERS? In the earlier phases of the study, teaching staff rated as sensitive and engaged using the Caregiver Interaction Scale (Arnett, 1989) worked in classrooms receiving ECERS scores indicating good or excellent global quality (Whitebook et al., 2001).

Second, we wanted to know if the ECERS-R remedied some of the weaknesses of the original ECERS related to cultural sensitivity and staff stability. For example, the item on cultural diversity and sensitivity in the ECERS-R (item 28) was expanded upon from the original

ECERS (item 30). Most notably, the newer version places more emphasis on adult behavior than the original, assessing programs for demonstrated prejudice as well as proactive intervention to counteract it, and promotion of understanding and acceptance of diversity.

To assess improvements in the revised edition and to determine how centers that scored well on the ECERS-R fared on other measures of culture and staff stability, we examined the following:

1. Are the items that focus on culture and diversity from the original and revised ECERS comparable? Do centers that score well on the “cultural awareness” item of the ECERS also score well on the “promoting acceptance of diversity” item of the ECERS-R?
2. Research suggests that non-English speaking children benefit from having a teacher who speaks their home language (Chang et al., 1996). Do ratings for centers that employ staff who speak the home language of children they care for differ from centers whose staff do not speak the home language of non-English speaking children?
3. We also examined the relationship between the ECERS-R and staff stability. Do centers with good ECERS-R scores have lower staff turnover, retain a greater percentage of highly-skilled teachers and pay staff better wages? In our earlier assessment, centers that retained a greater percentage of highly-skilled teachers and paid higher wages were significantly more likely to receive good or better ratings on overall classroom quality (Whitebook et al., 2001).

#### Method

## *Sample*

### *Selection of Centers*

Forty-three centers were selected for participation in this study. The centers were located in the Northern California counties of San Mateo, Santa Clara and Santa Cruz. These communities share certain features, including a mix of high-, middle-, and low-income neighborhoods. The majority of the centers operated in middle-income neighborhoods (65 percent) on a nonprofit basis (72 percent).

The sample was part of a larger longitudinal study on center-based quality focused on the NAEYC accreditation process and the impact of child care staffing problems on programs' abilities to improve and sustain quality (Whitebook et al., 2001). This sample included 19 NAEYC-accredited centers and 24 non-accredited centers. The latter group included both centers that had sought but failed to achieve accreditation and those that had not sought accreditation. The non-accredited centers were matched to the accredited centers by nonprofit or for-profit status (center auspices) and income level of the census tract in which they were located.

### *Selection of Classrooms*

A sample of 68 classrooms was observed in the 43 centers. Two classrooms serving preschool-age children in each center were randomly selected for observations. In centers that did not include two preschool classrooms because of center size and/or age distribution of children, only one classroom was observed. Two classrooms were observed in 25 centers, and a single observation occurred in 18 centers.

### *Selection of Participants*

*Teaching staff.* For the ECERS and ECERS-R measures we observed all teaching staff in each selected classroom. However, to collect more in-depth demographic information (e.g., languages spoken by teachers and children in their classroom), we chose to interview the head or lead teachers in each classroom, because such staff typically set the tone and style for classroom activities and interactions. If a classroom had co-teachers, both teachers were interviewed. If the classroom had an assistant teacher, she/he was selected to participate in an interview in order to capture the perspective on center processes based on differing roles. If more than one non-lead teacher or assistant worked in the same classroom, we used random sampling to select assistants or teachers. In the 43 observed classrooms, 117 teaching staff were observed and interviewed: 71 percent teachers, 17 percent assistants, and 12 percent teacher-directors.

*Directors.* We interviewed the director of each center to ensure that a person with an overview of center operations and access to center records could provide details about salaries, turnover and staff qualifications. Directors' job definitions varied depending on the size and structure of each center. In some cases, directors or assistant directors worked in the classroom along with performing administrative functions; in others, the director's role involved minimal classroom contact and focused primarily on administrative tasks.

### *Measures*

Measures included observational instruments routinely used to observe and assess child care center quality and teacher-child interaction as well as interview protocols for teaching staff and center directors adapted or developed for the study.

#### *Classroom Observations*

Two research assistants completed observational assessments in each classroom. One completed the original ECERS and the other completed the revised version of the ECERS

(Harms & Clifford, 1980; Harms et al., 1998). The 1980 version of the ECERS is a 37-item instrument that focuses on whether programs include developmentally appropriate materials, activities and interactions around seven content areas, as detailed in Table 1. The ECERS is scored on a seven-point scale: 7=excellent, 5=good, 3=minimally adequate and 1=inadequate care. Even-number scores indicate that some of the requirements of the higher rating are met, but others are not (e.g., a score of 2 indicates that the center is better than a score of 1 and meets some but not all the requirements for a score of 3).

The revised edition of the ECERS (ECERS-R), published in 1998, employs slightly different content areas than the ECERS, as shown in Table 1. Like the ECERS, scoring of the ECERS-R is on the same seven-point scale (7=excellent to 1=inadequate care). Unlike the original ECERS, however, a slightly different scoring convention is used for even-number scores on the ECERS-R. For the ECERS-R, even-number scores indicate that half or more (but not all) of the requirements of the higher rating are met. For both measures, we calculated global scores based on the average of all items.

To measure adult-child interactions, we selected the Caregiver Interaction Scale (CIS; Arnett, 1989), which in previous large-scale studies has been found to predict teachers' engagement with children and children's language development and security of attachment (Helburn, 1995; Howes, Phillips, & Whitebook, 1992). The CIS is a 26-item observational scale that rates an individual teacher's style of behavior. For this study, we used two subscales previously calculated in other studies (e.g., Helburn, 1995; Whitebook et al., 1990): (a) sensitivity (i.e., 10 items measuring a teacher's degree of warmth, attentiveness and engagement) and (b) style or harshness (i.e., 9 items measuring a teacher's level of punitive and critical interactions). Items are scored on a scale from 1 to 4. A score of 1 indicates that a given behavior

is “never true,” whereas a score of 4 indicates that the behavior is “often observed.” Higher scores for sensitivity and lower scores for harshness are therefore considered desirable. The CIS is used to rate a single teacher, in contrast to the ECERS and ECERS-R, which are used to rate an entire classroom and all teachers in that classroom. The CIS was used to rate the style of the head teacher (or two co-teachers, if the classroom had co-teachers) in each classroom. Subscales had acceptable Cronbach’s alpha of .90 for sensitivity and .82 for harshness.

### *Interviews*

Director and teacher interviews were adapted from measures used in the National Child Care Staffing Study (Whitebook et al., 1990). Directors provided information about the turnover, compensation and professional background of all staff employed at their centers. The head teacher and one assistant teacher in each classroom, or two co-teachers (if the classroom had co-teachers) provided information about center practices related to inclusion and diversity, and the linguistic match among children, their parents and staff.

### *Procedure*

The research assistant team was composed of six people with experience in the early childhood education field. Three research assistants were trained to complete the ECERS and three were trained to collect data on the ECERS-R. Four of the research assistants had no prior experience with either measure; two had collected ECERS-R data for this study and also had prior experience with the original ECERS. ECERS and ECERS-R training sessions were conducted separately. The entire research team was trained to conduct observations and interviews during a five-day training session followed by several practice visits. Inter-rater reliability was established to a criterion of 85-percent agreement within one rating point for all observational measures prior to data collection, and again, halfway through data collection.

Reliability was established between all six research assistants and two senior project researchers trained by Thelma Harms. Average weighted Kappas were .74 for the ECERS, .71 for the ECERS-R and .80 for the Caregiver Interaction Scale.

Following an initial phone call to directors, research assistants contacted the directors again by phone to make appointments to collect data at the centers. In each center, data collection began with a two-hour classroom observation. During the observation, both research assistants were in the classroom, one scoring the classroom on the ECERS, the other completing the ECERS-R. The researchers, as much as possible, stayed together, shadowing the same groups of teachers and children. Any questions for staff necessary to complete the ECERS and/or ECERS-R scoring were made in the presence of both research assistants. Unless the program only operated in the afternoon, all observations were conducted in the morning. If two classrooms were observed in one center, observations occurred over two days, scheduled consecutively whenever possible. In order to ensure that their scoring was not influenced by each other's perceptions of the classrooms, researchers were instructed not to discuss the observations.

Following the observation, the research assistants arranged to interview the teacher(s) about their own background, their language skills and the languages spoken by children in their classroom. These interviews took place generally during lunch or nap time, sometimes at the end of the day, or, if necessary, on another day. The director interviews occurred following the observations, typically in the afternoon. Every effort was made to accommodate the participants' schedules, with the exception of scheduling interviews prior to observations. Director interviews lasted an average of one-and-one-half to two hours. Teaching staff interviews lasted from half an hour to one hour. Although both research assistants were present to ask teachers questions necessary to complete the ECERS and ECERS-R, only one interviewed each teacher about her

own demographic characteristics and the linguistic characteristics of the children in her classroom.

### *Plan of Analyses*

Pearson correlations and descriptive analyses were used to compare global ECERS and ECERS-R scores for each classroom. We used principal components analysis to compare the factor structure of the ECERS and ECERS-R. Pearson correlations, chi-square analyses and t-tests were used to examine the relationship between child care quality and teacher-child interactions, and to examine how well the ECERS-R assessed two areas of growing concern to child care experts: cultural sensitivity and staff stability.

### Results

#### *Comparisons of ECERS and ECERS-R*

##### *Global Ratings*

Before 1998, most large-scale studies of center-based child care quality used the 1980 version of the ECERS as their outcome measure. To know whether data from studies conducted after 1998 using the ECERS-R can be fairly compared to data from studies using the 1980 ECERS, we conducted simultaneous assessments using both measures in a sample of 68 classrooms that were part of a larger study of child care quality, staffing and NAEYC accreditation. The distribution of scores on the ECERS and ECERS-R were remarkably similar, with 47 percent and 44 percent respectively of classrooms providing care rated as adequate and a little more than half (53 percent) of all classrooms rated in the good-to-excellent range on both the ECERS and the ECERS-R (see Table 2).

In our sample of 68 classrooms, within-classroom scores on the ECERS and ECERS-R were highly correlated ( $r(68) = .79, p < .001$ ). ECERS scores ranged from 2.65 to 6.38, a result

very similar, although not identical, to scores of 3.00 to 6.35 on the ECERS-R (see Table 3). One classroom had identical ECERS and ECERS-R scores. The mean absolute difference between ECERS and ECERS-R scores was .40 (SD=.33, n=68 classrooms). Forty percent of classrooms had absolute difference scores between the ECERS and ECERS-R of .25 or less; 34 percent had absolute difference scores between .26 and .50; and 26 percent had absolute difference scores greater than .50. More than half of the classrooms (60 percent) scored higher on the ECERS than on the ECERS-R, with differences in scores ranging from 0.01 to 1.44 points. Thirty-eight percent of classrooms had higher ECERS-R than ECERS scores, with differences ranging from 0.06 to 1.33 points.

#### *Subscales and Factor Structure*

The National Child Care Staffing Study was the first to perform factor analyses on the ECERS scale items (Whitebook et al., 1990). The first subscale, “appropriate caregiving,” captured items pertaining to adult-child interactions, supervision and discipline. The second subscale, “developmentally appropriate activity,” pertained to materials, schedule and activities (Whitebook et al., 1990). Subsequently, other researchers have found similar empirical dimensions using the scale (Helburn, 1995; Rossbach, Clifford, & Harms, 1991). The factors have been alternately labeled as “materials” and “tone” subscales (Whitebook et al., 1997). We were interested in whether the revised version of the ECERS would submit to the same factor structure as the ECERS. We performed exploratory factor analyses on items from the ECERS and ECERS-R.

We first ran a principal components analysis with a varimax rotation and loading values greater than .45, including all items of the ECERS except the items referring to adult provisions. The scree test pointed to solutions with two to nine factors, with a two-factor solution proving

the best fit, accounting for 45.43% of the variance. This two-factor solution demonstrates a similar pattern to the two subscales described by Whitebook et al. (1990), with the first factor including items pertaining to developmentally appropriate caregiving and interactions, or tone, and the second factor including items pertaining to developmentally appropriate activities and materials, or simply materials. To create more parsimonious factors, all items with overlapping factor loadings were deleted and re-analyzed following the same procedure. See Table 4 for final ECERS factor loadings.

To investigate whether the ECERS-R would also result in a similar two-factor solution representing similar items, we followed the same procedures described above for an exploratory principal component analysis on the ECERS-R data. The scree test pointed to solutions with two to six factors, also with a two-factor solution proving the best fit, accounting for 67.1% of the variance. Although a two-factor solution fits the ECERS-R data, the items included in each factor do not map as clearly onto the previously described subscales of tone and materials. While there is a general pattern of more items pertaining to materials grouping together in factor 1 and more items pertaining to caregiving and interactions grouping together in factor 2, there is reversal of items that are clearly differentiated in the current and previously described ECERS factor analyses. See Table 5 for final ECERS-R factor loadings.

Considering the expectation that the ECERS-R would factor into a similar two-factor solution as the original ECERS, we consulted with colleagues conducting larger studies using the ECERS-R. These personal communications suggest that with a larger sample a simpler two-factor solution generally representing “Teaching and Interactions” and “Provisions for Learning” will persist (personal communication with C. Howes, 2003).

We next ran a reliability analysis on just the items found to persist in the ECERS-R two-factor solution. An internal consistency estimate of reliability, coefficient alpha, was computed for the total scale and the two subscales, Teaching and Interactions and Provisions for Learning, based on the two-factor solution presented in personal communication with Howes (2003) for the ECERS-R. Two items, Free Play and Group time, were accounted for in both factors conducted on the ECERS-R. These two items were not included in the reliability analysis. Since the two subscales within the ECERS-R measure contain items that are interrelated and all describe different aspects of quality, a high reliability for the entire scale as well as for each of the subscales is consistent with the underlying constructs of the scale. As expected, the total scale had an alpha of .94.

The subscales of Teaching and Interactions and Provisions for Learning had alphas of .84 and .81 respectively. We next assessed the discriminant validity of the subscales to determine whether all or some of the items on each subscale strongly related to the constructs underlying them. Two new variables, which had the sum of the items for that subscale divided by the number of items, were created. For each subscale, items were correlated with their own scale (with the item removed) and with these sum variables. As shown in Table 6, all items except one in each subscale are correlated more strongly with their own scale than with the other scales. The Teaching and Interactions subscale included one item (Encouraging Children to Communicate) that was correlated more strongly with the Materials subscale scale than with its own subscale. Similarly, the Materials subscale included one item (Nature/Science) that was correlated more strongly with the Teaching and Interactions subscale than with its own subscale. Each of these items was removed from the subscale reliability analysis. The revised subscale alphas had

satisfactory alpha coefficients of .82 and .81 for Teaching and Interactions and Provisions for Learning, respectively.

#### *Global Scores and Teacher Behavior*

First we calculated Pearson correlations between the two global ECERS scores and the sensitivity and harshness subscales of the CIS (Arnett, 1989). Higher scores on the original and revised ECERS were associated with more sensitive (ECERS  $r(68) = .60, p < .001$ ; ECERS-R  $r(68) = .54, p < .001$ ) and less harsh teachers (ECERS  $r(68) = -.56, p < .001$ ; ECERS-R  $r(68) = -.52, p < .001$ ).

We next examined whether there were differences in teacher sensitivity and harshness in classrooms rated good or better on the ECERS and ECERS-R (global score of 5 or better) and classrooms rated mediocre or poor (global score less than 5). As shown in Table 7, both the ECERS and ECERS-R measures differentiated teaching staff behaviors. Classrooms considered high in quality (a score of 5 or above) on the ECERS and ECERS-R were more likely to have teachers who were more sensitive and less harsh.

#### *Assessing the Revisions*

*Culturally sensitive environments.* Classrooms in our sample served families who were ethnically and linguistically diverse. Teaching staff reported that on average, 39 percent of the children in their classrooms were children of color. Nearly half of the classrooms had children whose home language was not English. Slightly more than one-quarter of teaching staff (28 percent) in our sample reported that parents had difficulty communicating with staff at their center because of a language barrier. All of the classrooms with children who spoke English had teachers who also spoke English, but only one-third (34 percent) of classrooms with at least one child who spoke Chinese or Spanish employed a teacher who also spoke that language.

Disturbingly, in this sample of relatively high-quality programs serving such a diverse population, 70 percent of teachers reported little or no time to discuss cultural differences among staff during meetings, and 66 percent reported little or no time to discuss cultural differences between staff and families at staff meetings.

We began our analyses around cultural sensitivity by comparing scores on the Cultural Awareness and Promoting Acceptance of Diversity items in the original and revised measures. Classrooms, overall, scored very low on these items. Eighty-four percent of classrooms scored below a 5 on cultural diversity items on both the ECERS and ECERS-R. Only three classrooms scored in the good range on both the ECERS and ECERS-R. However, independent scoring of this item on the ECERS and ECERS-R appears consistent. Within-classroom scores on the Cultural Awareness item of the original ECERS and the Promoting Acceptance of Diversity item on the revision were significantly correlated ( $r(68) = .65, p < .001$ ).

Research suggests that children who speak languages other than English benefit from teachers who speak the same home language. By collecting language information on teachers and the children in their classrooms we were able to determine which classrooms had non-English speaking children and whether such classrooms also had teachers who could speak the same language. We next examined whether global ECERS and ECERS-R scores and cultural diversity individual item ratings differed in classrooms that met this language match for non-English speaking children. We found no differences between ECERS and ECERS-R global scores or the cultural diversity item scores in classrooms that did or did not have a language match between teachers and children. In addition, overall quality scores and cultural diversity item scores were no worse in classrooms in which teachers reported communication problems

with parents because of a language barrier or in which staff felt they had insufficient time to discuss issues related to culture among themselves, parents or children.

*ECERS and ECERS-R ratings and staff stability.* Previous research (Helburn, 1995; Phillips et al., 2000; Whitebook et al., 1990) demonstrated a relationship between staff stability as measured by rates of annual turnover and program quality as measured by the ECERS. In previous phases of the current study, centers that retained a greater percentage of highly skilled teachers were significantly more likely to receive good or better ratings on overall classroom quality (Whitebook et al., 2001).

Two measures of staff stability were employed in this study. Annual turnover was based on directors' reports of the number of teaching staff departures in the previous 12 months; four-year turnover rates were based on a count of the number of teaching staff working in the centers in 1996 no longer on staff at the time of our 2000 visit. Directors reported on average that 30 percent of teaching staff ( $SD=36$  percent, range=0 to 175 percent) had left in the past 12 months. Four-year turnover rates averaged 71 percent ( $SD=27$  percent, range=0 to 100 percent). We first calculated Pearson correlations between our measures of staff stability and global ECERS and ECERS-R scores. No correlations were significant. In addition, centers that scored in the good-to-excellent range on the ECERS or ECERS-R (scores of 5 or greater) were no more likely to have low turnover rates than were centers with scores considered adequate or poor (less than 5) on either measure (see Table 8).

Previous research found that the strongest predictor of whether a center could sustain ECERS ratings of 5 or greater over time was the presence of a greater percentage of highly skilled teaching staff (Whitebook et al., 2001). "Skilled teaching staff" was defined as the percentage of teaching staff with a bachelor's degree and/or advanced training in early childhood

education in the center. In this study, we found that there were no significant relationships between the percentage of skilled staff at a center with scores on either the ECERS or the ECERS-R. Centers with higher ECERS (or ECERS-R) scores did not have a higher percentage of skilled staff than centers with lower scores (see Table 8). Although we found a moderate correlation between head teacher wages and scores on both the original ECERS ( $r(68) = .30$ ,  $p < .05$ ) and the ECERS-R ( $r(68) = .27$ ,  $p < .05$ ), findings regarding wages of highly skilled staff were not comparable on the two measures. In this sample, classrooms rated good or better (score greater than 5) on the ECERS paid their head teachers more ( $M = \$16.69$  per hour,  $SD = 5.45$ ) than did lower-quality classrooms (ECERS score less than 5;  $M = \$14.12$  per hour,  $SD = 4.79$ ;  $t(64) = -2.03$ ,  $p < .05$  (two-tailed)). The same pattern of results was not found when comparing wages and center quality using the ECERS-R.

### Discussion

This investigation sought to provide evidence regarding the comparability of the original and revised Early Childhood Environment Rating Scales and to explore whether the revised measure addresses key weaknesses identified by practitioners in the field. The results presented here suggest that the original ECERS and ECERS-R can be viewed, as their authors intended, as comparable measures of quality. Based on simultaneous observations using both instruments, scores were highly correlated and similarly distributed. In addition, both measures produced similar results when compared to another widely used measure of teacher behavior, the Caregiver Interaction Scale. Although results of the principal components analysis on ECERS and ECERS-R items need replication on a larger scale, in general two factors were found for both measures. Because we did not conduct assessments of children for this study, we were unable to ascertain whether the two versions of the ECERS predict child development outcomes

similarly. Additional research comparing these measures to confirm the findings reported here and to explore the relationship of the ECERS-R to child outcomes would be useful. Such investigation is particularly warranted given the increasing use of the revised measure not only in research but as an assessment tool upon which licensing, funding and professional reward decisions are based (Burton et al., 2000).

The degree to which the ECERS-R improves upon the original ECERS is a more complicated and perhaps pressing matter. With respect to cultural sensitivity and promoting tolerance, the revised measure improves upon the original by delineating proactive teacher behaviors, and by infusing the entire measure with indicators and examples of appropriate practices in this area. On the one hand, among a sample of programs in which a high percentage fall short of meeting the linguistic needs of children and parents, it is encouraging that scores on the focused cultural item fell below good or excellent for the vast majority of programs. The fact that programs that fall short on linguistic match can still be rated as high in quality, however, is troubling. Those wishing to assess programs in terms of cultural sensitivity will need to supplement ECERS ratings with another form of assessment. California Tomorrow and others have been among a limited number devoted to understanding and measuring the role of culture and diversity in child care settings (Chang et al., 1996). To understand how culture shapes children's experiences, it seems necessary to look not only at quality, as measured by the ECERS-R, but also at the specific goals and intentions that the participants have as they engage in the observable activities. Indeed, Wishard, Shivers, Howes and Richie (in press) argue that child care practices are deeply rooted within an ethnic, cultural and historical community that shapes how and why activities are carried out. Future editions of the instrument should continue to grapple with these issues.

With respect to staff stability, the fact that centers rated high in quality by both the ECERS and ECERS-R did not differ from lower-quality programs in terms of turnover may be a function of several factors, some of which are not related to the measure itself. For example, a restricted range of turnover and/or quality may account for this finding, and there is evidence to suggest that both were true. Our sample included a very high percentage of high-quality programs. In addition, data were collected in 2000, at the height of an economic boom during which all centers (and many other low-paying industries) were struggling to recruit and retain staff. The communities in which centers in the sample were located were particularly hard hit because they were in the heart of the high-tech Silicon Valley, where high-paying entry level jobs were plentiful. Also, class size reduction in the elementary schools during this period was drawing many staff with bachelor's degrees into better-paying public school jobs.

It is also possible, however, that neither measure sufficiently captures staff stability, an organizational process that has been demonstrated to impact children's experience in child care settings (Helburn, 1995; Howes & Hamilton, 1993; Whitebook et al., 1990). Much of the emphasis in both versions of the ECERS focuses on material items or concrete practices that are relatively easy to implement and assess. It is far more difficult to assemble and sustain a consistent teaching workforce, and its absence or disruption, while difficult to assess, impacts children's development and behavior, as well as their experience in care (Whitebook & Bellm, 1999). Children experiencing high rates of turnover suffer in their language and social development, building vocabulary skills more slowly and exhibiting aggressive behavior toward their peers (Howes & Hamilton, 1993; Whitebook et al., 1990). Given the link between turnover and quality, the question remains why the authors of the ECERS-R did not include questions on staff stability in their revision. That these structural measures are not adequately measured by

the ECERS-R is not necessarily a critical flaw, but it is important to recognize that the ECERS-R cannot stand alone as a comprehensive measure of quality. ECERS-R assessments should be supplemented with other measures of quality, particularly one that accounts for the frequency and magnitude of changes among staff with varying levels of skill, to get a complete picture of the health and practices of a center.

An area of increasing interest to both researchers and practitioners is a program's sensitivity to special needs children. The ECERS-R takes positive steps in articulating good practice in this area. However, the fact that the most focused question on this topic is scored only if the program has a special needs child enrolled suggests that researchers or parents wishing to assess programs' potential to include children with special needs will need to rely on supplementary measures. In our study, only 19 classrooms included a special needs child. This small sample makes it difficult to empirically address all questions related to this item. Our data suggest, however, that teacher' assessments of appropriate environments and resources for special needs children did not necessarily mesh with their scores on the special needs item. For example, classrooms in which teachers assessed the environment as inappropriate or lacking support and resources did not receive lower global ECERS or ECERS-R scores than classrooms with more appropriate environments and supports for special needs children. It is possible, of course, that these teachers were judging their work harshly and could still be doing a very good job for the special needs children in their care. With respect to resources and supports, it is possible, but less likely, that their assessments of inadequate supports were unreasonable. Additional research in this area is needed.

The revision of the ECERS has implications for practice and policy as well as empirical research. The ECERS-R is increasingly being used both as a tool for self-assessment by teachers

and directors and as a measure of program assessment by funders and policy makers (Burton et al., 2000). When using the ECERS-R as a self-assessment tool, global scores may be of less importance than scores on individual items and indicators where teachers and directors can consider their content to determine the strengths and weaknesses of a classroom. Updating the ECERS to include more items of interest to quality experts and clarifying the indicators of each item has added to its value as a self-assessment tool. Unfortunately, self-assessment is prone to inherent bias that exists when teachers evaluate themselves or directors evaluate their own centers. Practitioners need to be mindful of this bias and ensure that evaluators are trained on the instrument so that the quality of programs is not overestimated.

Although evidence to date is anecdotal, those using the ECERS as an assessment tool also express concerns regarding its sensitivity to underlying dynamics that drive quality. Some have suggested that as the ECERS is increasingly used as a self-assessment tool, we will witness a type of grade or score inflation in measuring quality. Centers will purchase the equipment and materials linked to high scores on the ECERS-R, and implement the concrete procedures, such as those focused on hygiene, that are clearly laid out in the measure. These are positive practices, and implementing them will raise scores and improve key aspects of program quality. But higher ratings may camouflage troubling staff behaviors and problems in the adult work environment that make it impossible to develop and sustain the good practices for children and that demand more depth of understanding of child development and exposure to skilled role models. This has special implications for those who use results on the ECERS-R to make major funding and policy decisions. Other measures in addition to the ECERS-R must be used to provide a complete picture of center practices that include cultural issues, troubling staff behaviors, and problems in the adult work environment that contribute to center quality.

Both versions of the ECERS have helped to stimulate discussions of quality and efforts to improve it. Yet in this era when concrete outcome measures are increasingly heralded as key to educational reform, there remains an elusive aspect to assessing quality, whether of early childhood programs or K-12 classrooms, or even assessing which students will succeed in college. Tools such as the ECERS are a good starting point for measuring progress in program improvement. The ECERS-R furthers the discussion of what constitutes child care quality, and provides one approach to measuring it. But we still have more to learn about how to measure quality, and most importantly, how we can work to improve teachers' ability to recognize it and implement it for children and families. Let the conversation continue.

## References

- Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology, 10*, 541-552.
- Bellm, D., Burton, A., Shukla, R., & Whitebook, M. (1997). *Making work pay in the child care industry: Promising practices for improving compensation*. Washington, DC: Center for the Child Care Workforce.
- Burton, A., Mihaly, J., Kagiwada, J., & Whitebook, M. (2000). *The CARES Initiative in California: Pursuing public policy to build a skilled and stable workforce, 1997-2000*. Washington, DC: Center for the Child Care Workforce.
- California Early Childhood Mentor Program. (2001). *Annual report, 2000-2001*. San Francisco: California Early Childhood Mentor Program.
- Chang, H., Muckelroy, A., & Pulido-Tobiassen, D. (1996). *Looking in, looking out: Redefining child care and early education in a diverse society*. San Francisco: California Tomorrow.
- Coelen, C., Glantz, F., & Calore, D. (1979). *Day care centers in the U.S.: A national profile 1976-1977*. Cambridge, MA: Abt Associates.
- Harms, T., & Clifford, R. (1989). *Family Day Care Rating Scale*, New York: Teachers College Press.
- Harms, T., & Clifford, R. (1980). *Early Childhood Environment Rating Scale*. New York: Teachers College Press.
- Harms, T., Clifford, R., & Cryer, D. (1998). *Early Childhood Environment Rating Scale, revised edition*. New York: Teachers College Press.

- Harms, T., Cryer, D., & Clifford, R. (1990). *Infant-Toddler Environment Rating Scale*. New York: Teachers College Press.
- Helburn, S. W. (Ed.). (1995). *Cost, quality and child outcomes in child care centers. Technical report*. Denver, CO: University of Colorado at Denver, Department of Economics, Center for Research in Economic and Social Policy.
- Holloway, S., Kagan, S., Fuller, B., Tsou, L., & Carroll, J. (2001). Assessing child care quality with a telephone interview. *Early Childhood Research Quarterly, 16*(2), 165-190.
- Howes, C. (2003). [Assessment of child care quality]. Unpublished raw data.
- Howes, C., & Hamilton, C. (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., Phillips, D. A., & Whitebook, M. (1992). Thresholds of quality in child care centers and children's social and emotional development. *Child Development, 63*, 449-460.
- Keyserling, M. (1972). *Windows on day care: A report on findings of the National Council of Jewish Women*. New York: National Council of Jewish Women.
- Lamb, M. E. (1998). Nonparental child care: Context, quality, correlates. In W. Damon, I. E. Sigel & K. A. Renninger (Eds.), *Handbook of child psychology, Vol. 4. Child psychology in practice* (5th ed., pp. 73-134). New York: Wiley.
- Love, J., Schochet, P., & Mechstroth, A. (1996). *Are they in any real danger? What research does - and doesn't - tell you about child care quality and children's well-being*. Princeton, NJ: Mathematica Policy Research, Inc.

National Research Council and Institute of Medicine. (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press.

North Carolina Department of Social Services. (1999). *North Carolina child care regulations, section 2803*. Raleigh, NC: North Carolina Department of Social Services.

Phillips, D., & Howes, C. (1987). Indicators of quality in child care: Review of research. In D. Phillips (Ed.), *Quality in child care: What does research tell us?* (pp. 1-20). Washington, DC: National Association for the Education of Young Children.

Phillips, D., Mekos, D., Scarr, S., McCartney, K., & Abbott-Shin, M. (2000). Within and beyond the classroom door: Assessing quality in child care centers. *Early Childhood Research Quarterly, 15*(4), 475-496.

Rossbach, H. G., Clifford, R. M., & Harms, T. (1991). *Dimensions of learning environments; Cross-national validation of the Early Childhood Environment Rating Scale*. Paper presented at the American Educational Research Association, Chicago.

Whitebook, M., & Bellm, D. (1999). *Taking on turnover: An action guide for child care center teachers and directors*. Washington, DC: Center for the Child Care Workforce.

Whitebook, M., & Eichberg, A. (2002). Finding the better way: Defining and assessing child care compensation initiatives. *Young Children, 57*(3), 66-72.

Whitebook, M., Howes, C., & Phillips, D. (1990). *The national child care staffing study. Final report: Who cares? Child care teachers and the quality of care in America*. Washington, DC: Center for the Child Care Workforce.

Whitebook, M., Sakai, L., & Howes, C. (1997). *NAEYC accreditation as a strategy for improving child care quality: An assessment. Final report.* Washington, DC: Center for the Child Care Workforce.

Whitebook, M., Sakai, L., Gerber, E., & Howes, C. (2001). *Then and now: Changes in child care staffing, 1994-2000.* Washington, DC: Center for the Child Care Workforce.

Wishard, A., Shivers, E., Howes, C., & Richie, S., (in press). Child care program and teacher practices: Associations with quality and children's experiences. *Early Childhood Research Quarterly.*

Table 1

*Comparison of ECERS and ECERS-R Subscales and Items<sup>a</sup>*

ECERS	ECERS-R
<u>Personal Care</u>	<u>Personal Care Routines</u>
(1) Greeting/departing	(9) Greeting/departing
(2) Meals/snacks	(10) Meals/snacks
(3) Nap/rest	(11) Nap/rest
(4) Diapering/toileting	(12) Toileting/diapering
(5) Personal grooming	(13) Health practices
	(14) Safety practices
<u>Furnishings and Display for Children</u>	<u>Space and Furnishings</u>
(6) For routine care	(1) Indoor space
(7) For learning activities	(2) Furniture for care, play and learning
(8) For relaxation and comfort	(3) Furnishings for relaxation
(9) Room arrangement	(4) Room arrangement
(10) Child related display	(5) Space for privacy
	(6) Child-related display
	(7) Space for gross motor
	(8) Gross motor equipment
<u>Language-Reasoning Experiences</u>	<u>Language-Reasoning</u>
(11) Understanding of language	(15) Books and pictures
(12) Using language	(16) Encouraging children to communicate
(13) Using learning concepts	(17) Reasoning skills

(14) Informal use of language

(18) Informal use of language

Fine and Gross Motor Activities

Activities

(15) Perceptual/fine motor

(19) Fine motor

(16) Supervision of fine motor activities

(20) Art

(17) Space for gross motor

(21) Music/movement

(18) Gross motor equipment

(22) Blocks

(19) Time for gross motor activities

(23) Sand/water

(20) Supervision of gross motor activities

(24) Dramatic play

Creative Activities

(25) Nature/science

(21) Art

(26) Math/number

(22) Music/movement

(27) Use of TV, video, and/or computers

(23) Blocks

(28) Promoting acceptance of diversity

(24) Sand/water

(25) Dramatic play

(26) Schedule

(27) Supervision of creative activities

Social Development

Program Structure

(28) Space to be alone

(34) Schedule

(29) Free play

(35) Free play

(30) Group time

(36) Group time

(31) Cultural awareness

(37) Provisions for children with disabilities

(32) Tone

(33) Provisions for exceptional children

Interaction

(29) Supervision of gross motor activities

(30) General supervision of children

(31) Discipline

(32) Staff-child interactions

(33) Interactions among children

Adult Needs

(34) Adult personal area

(35) Opportunities for professional growth

(36) Adult meeting area

(37) Provisions for parents

Parents and Staff

(38) Provisions for parents

(39) Provisions for personal staff needs

(40) Provisions for professional needs of staff

(41) Staff interaction and cooperation

(42) Supervision and evaluation of staff

(43) Opportunities for professional growth

---

<sup>a</sup> Number in parenthesis indicates item number for that measure.

Table 2

*Distribution of ECERS and ECERS-R Scores for 68 Classrooms*

Percentage of classrooms with scores rated	Inadequate care 1.00-2.99	Adequate care 3.00-4.99	Good to excellent care 5.00-6.99
ECERS	0	47 percent	53 percent
ECERS-R	3 percent	44 percent	53 percent

Table 3

*Descriptive Results: ECERS and ECERS-R*

Measure	Mean	SD	Range
ECERS	4.91	0.72	2.65-6.38
ECERS-R	4.87	0.84	3.00-6.35

Table 4

*ECERS Principal Components Analysis*

ECERS Items	Tone	Materials
(16) Supervision of fine motor activities	.87	
(32) Tone	.85	
(14) Informal use of language	.81	
(13) Using learning concepts	.76	
(12) Using language	.76	
(26) Schedule (creative)	.71	
(29) Free play	.70	
(27) Supervision of creative activities	.68	
(1) Greeting/departure	.67	
(20) Supervision of gross motor activities	.62	
(30) Group time	.62	
(22) Music/movement	.59	
(31) Cultural Awareness	.54	
(8) Furnishings for relaxation and comfort		.80
(24) Sand/water		.64
(25) Dramatic play		.64
(6) Furnishings for routine		.63
(23) Blocks		.63
(33) Provisions for exceptional children		.61
(10) Child-related display		.59

---

(17) Space for gross motor	.56
(28) Space to be alone	.52
(4) Diapering/toilet	.49
(15) Perceptual/fine motor	.46
(7) Furnishings for learning activities	.42
(3) Nap/rest	.40

---

Table 5

*ECERS-R Principal Components Analysis*

ECERS-R Items	Factor 1	Factor 2
(24) Dramatic play	.87	
(10) Meals/ snacks	.83	
(23) Sand/water	.79	
(4) Room arrangement	.78	
(30) General supervision of children	.78	
(28) Promoting acceptance of diversity	.77	
(14) Safety practices	.73	
(15) Books and pictures	.73	
(12) Toileting/diapering	.70	
(32) Staff-child interactions	.67	
(34) Schedule	.64	
(22) Blocks	.59	
(8) Gross motor equipment	.53	
(36) Group time	.52	
(13) Health practices	.50	
(1) Indoor space	.49	
(11) Nap/rest		
(35) Free play		.82
(18) Informal use of language		.79
(17) Reasoning skills		.75

---

(33) Interactions among children	.75
(19) Fine motor	.73
(21) Music/movement	.72
(26) Math/numbers	.69
(2) Furniture for care, play and learning	.59
(9) Greeting/departing	.57
(7) Space for gross motor	-.52
(3) Furnishings for relaxation	.51
(29) Supervision of gross motor activities	
(5) Space for privacy	
(27) Use of TV, video, and/or computers	

---

Table 6

*ECERS-R Subscale Reliability Correlations*

Items	Provisions for Learning	Teaching & Interactions
(9) Greeting/departing	.37	<b>.54</b>
(16) Encouraging children to communicate*	.66	<b>.59</b>
(17) Using language to develop reasoning skills	.17	<b>.23</b>
(18) Informal use of language	.41	<b>.70</b>
(29) Supervision of gross motor activities	.14	<b>.44</b>
(30) General supervision of children	.47	<b>.60</b>
(31) Discipline	.63	<b>.74</b>
(32) Staff-child interactions	.42	<b>.58</b>
(33) Interactions among children	.25	<b>.55</b>
(4) Room arrangement	<b>.56</b>	.48
(5) Space for privacy	<b>.44</b>	.39
(8) Gross motor equipment	<b>.32</b>	.19
(19) Fine motor	<b>.43</b>	.24
(20) Art	<b>.60</b>	.39
(22) Blocks	<b>.64</b>	.33
(23) Sand/water	<b>.69</b>	.56
(24) Dramatic play	<b>.69</b>	.44
(25) Nature/science*	<b>.34</b>	.35
(34) Schedule	<b>.30</b>	.29

Correlations of each item with its own subscale after removing focal item in **bold**.

\* Item deleted from subscale

Table 7

*ECERS and ECERS-R Global Quality Ratings and Teacher Behaviors*

	ECERS score less than 5	ECERS score greater than or equal to 5	<i>t</i>	<i>df</i>	ECERS- R score less than 5	ECERS- R score greater than or equal to 5	<i>t</i>	<i>df</i>
Sensitivity	3.27	3.72	-4.12 <sup>+</sup>	42	3.21	3.78	-5.49 <sup>+</sup>	41
Harshness	1.57	1.27	3.71 <sup>**</sup>	37	1.57	1.28	3.51 <sup>**</sup>	37

<sup>\*\*</sup>  $p < .01$ ; <sup>+</sup>  $p < .001$  (two-tailed)

Table 8

*ECERS and ECERS-R Ratings: Average Staff Turnover and Percentage of Skilled Staff*

	Average percentage turnover, 4-year period; all teaching staff	Average director-reported 12-month teaching staff turnover	Average percentage of teaching staff with a bachelor's degree and/or advanced training in early childhood education
ECERS less than 5.0	73%	35%	33%
ECERS greater than 5.0	64%	21%	46%
ECERS-R less than 5.0	69%	30%	37%
ECERS-R greater than 5.0	66%	25%	43%