

## **Reliability and Validity of the Early Childhood Environment Rating Scale**

Richard M. Clifford

Stephanie S. Reszka

FPG Child Development Institute

University of North Carolina at Chapel Hill, USA

Hans-Guenther Rossbach

University of Bamberg

Germany

Copyright © 2010 by Richard M. Clifford, Stephanie S. Reszka & Hans-Guenther Rossbach.

This is a working paper and revisions and additions are made on an occasional basis. This is the January 2010 version of the paper. Permission is hereby given to photocopy this document for individual use. For permission to make multiple copies, contact Richard Clifford at [dickclifford.unc.edu](mailto:dickclifford.unc.edu).

## Uses and History of the ECERS and ECERS-R

There are four environment rating scales, each designed for a different segment of the early childhood field: Infants and Toddler Environment Rating Scale – Revised (ITERS-R), Family Child Care Environment Rating Scale – Revised (FCCERS-R), School Aged Environment Rating Scale (SACERS), and the Early Childhood Environment Rating Scale – Revised (ECERS-R). Each scale contains items assessing the physical environment, basic care of children, curriculum, interactions, schedule and program structure, and parent and staff needs.

The Environment Rating Scales were developed to evaluate the process quality in settings for children. Process quality refers to the experience of children within the care environment including their interactions with others, materials, and activities (Phillipsen, Burchinal, Howes, & Cryer, 1997). Process quality is assessed primarily through observation and has been found to be more predictive of child outcomes than structural indicators of quality such as staff to child ratio, group size, cost of care, and type of care (Whitebook, 1989).

High quality care environments for children must provide for three basic needs that all children have: protection of their health and safety, the facilitation of building positive relationships, and opportunities for stimulation and learning from experience (CITE website). All three components must exist to create a high quality environment, and these components are assessed by items on the Environment Rating Scales.

The most widely used of the Environment Rating Scales is the ECERS-R. The original ECERS (Harms and Clifford, 1980), the first of the environment rating scales, was designed to assist practitioners in examining early childhood environments in order to make improvements in the provisions for young children. To meet the demand of this kind of task, the instrument first had to be both theoretically and practically grounded. Conceptually the original ECERS items

are organized into seven subscales that guide the observer to practically meaningful areas of interest in early childhood classrooms. These include 1) Personal Care Routines, 2) Furnishings & Display for Children, 3) Language-Reasoning Experiences, 4) Fine and Gross Motor Activities, 5) Creative Activities, 6) Social Development, and 7) Adult Needs.

A revised version of the ECERS was released in 1998. The ECERS-R contains seven subscales including 1) Space and Furnishings, 2) Personal Care Routines, 3) Language-Reasoning, 4) Activities, 5) Interaction, 6) Program Structure and 7) Parents and Staff. The revisions of the original scale reflect changes that occurred in the early childhood field in the 18 years since the original ECERS was developed. The ECERS-R places greater emphasis on important and emerging issues in early childhood childcare such as the inclusion of children with disabilities, family concerns, and cultural diversity.

One additional change in the transition from the ECERS to the ECERS-R is the use of a more strict indicator system supporting the ratings using the revised scale. Using the ECERS-R, raters have less freedom in assigning scores given the stricter descriptions of the scores in the revised version. Studies comparing the rating scales in terms of whether or not this change leads to a decrease in the mean quality level (i.e., if differences between ratings with the ECERS and the ECERS-R are due to real quality differences or to the differences in applying the rating scales) show mixed results (R. Clifford & H.-G. Rossbach, in press; Clifford, 2005). The ECERS-R ratings may lead to a systematic decrease of the assessed quality level by about half a scale point in some samples. However, this difference was only found in German samples, whereas in the US the means seem not to be effected by the different methodological features of the rating scales (R. Clifford & H.-G. Rossbach, in press; Clifford, 2005).

This paper will explore the reliability and validity evidence currently available for the ECERS-R. In some cases there is little information available for the revised version, so the original ECERS will be cited instead.

### *Research Use*

The Environment Rating Scales have a long history of use in research projects. The original ECERS was used in large, national studies such as the Head Start FACES study which included over 400 classrooms in the United States. In addition, the ECERS and ITERS were used as the comprehensive quality measures in the National Child Care Staffing Study (Whitebook, 1989) and the Cost, Quality, and Child Outcomes Study (1995), the major studies of their time.

The ECERS-R is currently being used in major studies including the Early Head Start Study, Head Start Family and Child Experiences Survey (FACES 2003), Georgia Early Care Study (GECS), More at Four (MAF) Evaluation in North Carolina, National Center for Early Development and Learning (NCEDL), Study of Early Child Care and Youth Development (NICHD SECCYD), and the Preschool Curriculum Evaluation Research Program (PCER).

### *Program Evaluation and Improvement*

In addition of their use in research, the scales are used in a variety of ways including for self-assessment by center staff, preparation for accreditation, and voluntary improvement efforts by licensing or other agencies. For example, in the United States:

- Consultation in Connecticut for the inclusion of children with disabilities, in Colorado for the inclusion of children from ethnically diverse and financially impoverished areas
- Technical assistance tied to federal funding for improving child care programs in Arkansas

- Self-assessments in North Carolina's Smart Start program before programs can apply for individual grant funding
- Statewide Star Rated/Tiered License systems in North Carolina, Tennessee, and Oklahoma
- Other quality evaluation and improvement programs in California, Massachusetts, Montana, Mississippi, Kansas, Oregon, Kentucky, New Mexico, Georgia, Florida, Wisconsin, and Nebraska, as well as in Washington, D.C. and U.S. Military service sponsored care facilities

#### *International uses of ECERS*

In addition to the uses in the U.S., the Environment Rating Scales have been used in research studies and program improvement efforts in many other countries including Canada, Germany, Italy, Sweden, Russia, Iceland, Portugal, England, Spain, Austria, Singapore, Korea, Hungary and Greece. Example international uses of the Environment Rating Scales include:

- In Canada, the scales are available in both English and French for licensing and consultation
- In Sweden, several projects are using preschool teachers as leaders in program improvement efforts with the Swedish ECERS
- In Germany, the scales are presently being used by individual cities to evaluate the quality of child care and kindergarten programs

The Environment Rating Scales, in translation or with minor adaptations, have been shown to produce reliable and valid ratings in each country and region. In England, Greece, Germany, Portugal, Spain, and Austria, higher scores on the Environment Rating Scales have been shown to be related to more positive child development outcomes (Melhuish &

Petrogiannis, 1996). In spite of the cultural differences between these areas, each adheres to a core set of child development goals and early childhood practices that align with those assessed by the Environment Rating Scales (R. Clifford & H.-G. Rossbach, in press; Clifford, 2005).

The global quality of the child care environments assessed by the Environment Rating Scales can be measured meaningfully and with confidence across cultures (R. Clifford & H.-G. Rossbach, in press; Clifford, 2005; Peisner-Feinberg, et al., 2001; Zill & Resnick, 1998). The goal of the Environment Rating Scales is to provide a meaningful, stable, and reliable measure of global quality in environments for children. The main focus of this paper will be on the reliability and validity of the ECERS-R.

#### Reliability of the Environment Rating Scales

All of the Environment Rating Scales were developed in close collaboration with realistic field-based sites, and have been used in numerous research studies that assess the reliability of the scores. The reliability of an instrument refers to the consistency of scores over time and among different raters (Kubiszyn & Borich, 2003). There are several important types of reliability that are relevant to the Environment Rating Scales including test-retest reliability, internal consistency, and interrater reliability.

##### *Test-retest reliability*

Test-retest reliability is used as an indicator of the stability of scores on an instrument over time (Kubiszyn & Borich, 2003). Results across studies using the ECERS-R, indicate that the assessment of the global quality of an early childhood care and education setting as measured by the ECERS-R is stable over moderately long periods of time during a given school year where the teacher is stable in the classroom (R. Clifford & H.-G. Rossbach, in press; Clifford, 2005).

This finding does not address whether changes can occur with intervention, simply that left to the normal progress of the program the ECERS and ECERS-R scores remain stable.

Adaptations of the ECERS and ECERS-R for other languages and countries have demonstrated reliability as well. Using the German adaptation, the KES-R, quality was assessed at two time points ranging from 1 to 10 weeks apart with no intervention (R. Clifford & H.-G. Rossbach, in press; Tietze, Schuster, Grenner, & Rossbach, 2001). Included were 10 classes where the same observers applied the KES-R at the two measurement points and 10 classes where different observers used the KES-R. When the same observers were used, exact agreement of scores was reached on 73% of the items and agreement within one point was reached on 92% of the items.

*Reliability: Interrater reliability of the ECERS-R*

Interrater reliability refers to the agreement of independent raters when assigning scores (cite?). Field testing of the ECERS-R suggests that the ECERS-R demonstrates good interrater reliability at the indicator, item, and total scale levels. The percentage of agreement across all 470 indicators was 86.1%, with all indicators having a percentage of agreement over 70% (Harms, Clifford, & Cryer, 1998). At the item level, the percentage of exact agreement was 48%, with a percentage of agreement within one point of 71%. Finally, for the total score, the Pearson product moment correlation was .921 and the Spearman rank order correlation was .865. The interclass correlation for the total score was .915 (Harms, et al., 1998).

A study using the German version of the ECERS-R showed similar, if not more desirable, results across raters. When different observers were used, exact agreement was reached on 65% of the items and agreement within one scale point on 92% of the items. The Spearman rank order correlation was .92 using the same observers and .88 using different observers. These

results suggest that the German adaptation of the ECERS-R demonstrates both a high test-retest-reliability and a high stability of quality scores across time (R. Clifford & H.-G. Rossbach, in press; Clifford, 2005; Tietze, et al., 2001).

Extensive training, as well as follow-up reliability checks are important for obtaining reliable scores across multiple raters and time points. Studies often use raters who are previously trained to reach an 85% agreement level within one point across various settings (Cassidy, Hestenes, Hegde, Hestenes, & Mims, 2005). Reliability checks for the Cassidy and colleagues study (2005) were conducted every 6<sup>th</sup> observation, or for observers maintaining an agreement within one point at or above 90%, every 10<sup>th</sup> observation.

The authors of the Environment Rating Scales offer three or five day intensive workshops to help familiarize individuals with the use of the Environment Rating Scales. Additionally, video training materials are available for each scale, and resource guides are available for the ECERS-R and ITERS-R with in-depth information about each item and indicator. While there is no certification available for the use of these scales, it is important that users have a complete understanding of the scoring system and meaning of each indicator in order to reliably complete the scoring system.

#### *Reliability: Internal consistency of the ECERS-R*

The internal consistency of an instrument refers to the ability of scores from the instrument to provide a measure a single concept (Kubiszyn & Borich, 2003), in this case, the global quality of an environment.

The internal consistency of the Environment Rating Scales is generally assessed at the subscale and total scale level. Each subscale of the scales is intended to measure a specific aspect of quality, while the total scale is an indicator of the global quality of an environment. As



shown in TABLE ##, the subscale internal consistency scores for the ECERS-R range from .71 to .88, while the total scale internal consistency is .92, according to field tests of the instrument (Harms, et al., 1998).

TABLE ##: Intra-class correlations for ECERS-R subscales (Harms, et al., 1998)

Scale	Interrater Internal Consistency
Space and Furnishings	.76
Personal Care Routines	.72
Language-Reasoning	.83
Activities	.88
Interaction	.86
Program Structure	.77
Parents and Staff	.71
Total	.92

### Validity of the ECERS-R

In the most basic sense, validity is an indicator of whether the instrument measures what it is intended to measure (Kubiszyn & Borich, 2003). Validation of an instrument requires gathering evidence that support the inferences to be made based on the scores obtained from the assessment (Crocker & Algina, 1986).

Just as there are several forms of reliability, there are multiple indicators of the validity of scores on an assessment. Those relevant to the Environment Rating Scales include content validity, predictive validity, and concurrent validity.

#### *Content validity*

Content validity is “the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose”

(Haynes, Richard, & Kubany, 1995, p. 237). In other words, content validity is an indicator of how well the items of the instrument represent the overall domain of interest (Bailey, 2004).

Crocker and Algina (1986) outline the following steps used to establish content validity: a) defining the performance domain; b) selecting a panel of qualified experts in the content domain; c) providing a structure to allow matching of items to the domain; and d) collecting and summarizing data (p. 219). During the development of the original *ECERS*, the authors contacted seven nationally recognized experts in day care and early childhood (Harms & Clifford, 1980). These experts rated the importance of each item in early childhood programs. Overall, 78 percent of the items were rated as of high importance. The authors then made minor modifications to the scale, which should increase the validity (Harms & Clifford, 1983).

### *Predictive Validity*

Predictive validity, like concurrent validity, is a subset of criterion-related validity, scores are predictive of future scores or events (Bailey, 2004). Predictive validity is perhaps the most thoroughly researched form of validity related to the *ECERS-R*. REFERENCE CHART IN APPENDIX

Research suggests that there is a link between the quality of children's early care environments and their academic and cognitive outcomes (Sammons, et al., 2003a). Higher quality scores are related to children's cognitive development as assessed by the Bayley Scales of Infant Development – Mental Development Index (Love, et al., 2004).

Math/Number. Specifically, research findings indicate that there is a positive relationship between the social interaction subscale on the *ECERS-R* and children's early number concept development (Sammons, et al., 2003a). Additionally, higher quality scores on a shortened version of the *ECERS* are associated with higher scores on the *Woodcock-Johnson-R* (Woodcock

& Johnson, 1990) math achievement applied problems subset (Peisner-Feinberg, et al., 2001).

Furthermore, the Teaching and Interactions factor of the ECERS-R is related to children's performance on the Woodcock-Johnson-R math achievement applied problem subset during pre-kindergarten and kindergarten (Burchinal, et al., 2008).

Language/literacy. Scores on the Environment Rating Scales have been shown to be predictive of children's language and literacy performance. Specifically, higher quality is associated with children's development of receptive language, print awareness, and book knowledge. Children in higher quality environments as assessed by the ECERS-R and ITERS tend to display higher scores overall on the PPVT-III for children in Early Head Start (Love, et al., 2004). Similarly, other studies have noted an association between higher scores on a shortened version of the ECERS and children's scores on the PPVT-R, indicating greater receptive language abilities (Peisner-Feinberg, et al., 2001). Children's print awareness scores on Concepts About Print (Zill & Resnick, 1998) and greater book knowledge are associated with higher quality in Smart Start preschool as assessed by the ECERS (Bryant, et al., 2003). Ratings on the ECERS-R have been shown to be related to children's expressive language development in prekindergarten (Mashburn, et al., 2008). Additionally, the Teaching and Interactions factor is related to children's expressive language scores on the *Oral Expression Scale (OWLS)* and receptive language scores on the *Peabody Picture Vocabulary Test (PPVT-III)* (Burchinal, et al., 2008).

Social outcomes. Several important social outcomes have been shown to be related to ECERS-R scores. Specifically, there is a positive relationship between scores on the social interaction subscale and children's scores on a measure of independence, concentration, cooperation, and conformity skills in preschool (Sammons, et al., 2003b). Additionally, there is

a positive relationship between classroom scores on the language and reasoning subscale and children's cooperation and conformity skills (Sammons, et al., 2003b). Finally, research shows a negative relationship between the space and furnishings subscale and children's anti-social/worried behaviors (Sammons, et al., 2003b). Other studies have found a relationship between the total score on the ECERS-R and children's socio-emotional development (Montes, Hightower, Brugger, & Moustafa, 2005). Using the ECERS-R and the *Teacher-Child Rating Scale (T-CRS 2.1)* (Hightower, et al., 1986) in urban centers serving low-income children, Montes and colleagues (2005) found that higher quality classrooms were associated with a significant decrease in socio-emotional risk factors for children (e.g., lack of behavior control, poor social skills). Also using the *Teacher-Child Rating Scale*, there is a relationship between the Teaching and Interactions factor of the ECERS-R and children's social competence scores (a composite including assertiveness, frustration tolerance, task orientation, and peer social skills) (Burchinal, et al., 2008).

#### *Concurrent Validity*

Concurrent validity is subset of criterion-related validity, correlation to scores from another instrument (Bailey, 2004). Specifically, criterion-related validity is a description of the relationship between scores and the measurement of the criterion when they are made at the same time (Crocker & Algina, 1986, p. 224).

During the creation of the original ECERS, in addition to eliciting feedback from experts, the authors assessed the ease of training observers to use the ECERS as an indicator of the validity (Harms & Clifford, 1983). Two sets of observers, one group with experience in the early childhood field, and another with little or no background in child development, were trained to use the ECERS. The observers were paired, one from each group, to assess the quality

of 18 classrooms independently. –ASK Dick about the Berns (1979) study... extra group of raters

One method of establishing concurrent validity is to compare the correlation of scores on a measure with those of a previously validated measure that was designed to measure the same construct.

“Concurrent validity is studied when one test is proposed as a substitute for another (for example, when a multiple-choice form of spelling test is substituted for taking dictation), or a test is shown to correlate with some contemporary criterion (e.g., psychiatric diagnosis)” P. 282 (Cronbach & Meehl, 1955)

There are several instruments available that measure aspects of child care environments. Examples include the CLASS, Snaphot, CIS, ELLCO, ECERS-E. Correlations of subscales, total scores, and factors of the ECERS-R and these scales is provided in Appendix XX.

The CLassroom Assessment Scoring System (CLASS) (Pianta, La Paro, & Hamre, 2008) examines adult-child interactions in the classroom environment as an indicator of the quality of the environment (Pianta, et al., 2008). This rating scale has three broad domains related to interactions in the classroom: Emotional Climate (teacher and child interaction qualities, peer interactions, sensitivity, responsiveness, and secure base provided by teacher), Management (over-control, flexibility for child interests, behavior management, monitoring and dealing with behaviors, routines, productivity), Instructional Support (support for concept development, formats, engagement, quality of feedback, encouraging engagement) (La Paro, Pianta, & Stuhlman, 2004).

The *Emerging Academics Snapshot* (Ritchie, Howes, Kraft-Sayre, & Weister, 2001) is a time sampling measure of children’s engagement. Each child is observed in 20-second interval “snapshots,” followed by a 40-second coding period. The *Snapshot* contains codes for activity settings (i.e., basics, free choice, individual time, meals, small group, and whole group), pre-

academic engagements (i.e., aesthetics, fine motor skills, gross motor skills, letter/sound, mathematics, oral language development, pre-reading, read to, science, social studies, and writing), and teacher-child interactions (i.e., routine, minimal, simple, elaborated, scaffolding, and didactic).

The *ELLCO*, or *Early Language & Literacy Classroom Observation Toolkit* (Smith, Dickinson, Sangeorge, & Anastasopoulos, 2002) is designed to provide a measure of the academic environment of young children. The primary focus of the *ELLCO* is to assess the support and resources available to facilitate the language and literacy development of children. It consists of a *Literacy Environment Checklist (LEC)* that requires observation of the materials (books and writing materials) and use of these materials in the classroom, a *Classroom Observation and Teacher Interview (COTI)* related to areas such as general classroom organization, climate, management, and support for literacy, and a *Literacy Activities Rating Scale (LARS)* to assess book reading and writing (Smith, et al., 2002).

The *Caregiver Interaction Scale*, or *CIS* (Arnett, 1989) assesses the relationship between caregivers and children based on observations of caregivers' behaviors. The *CIS* consists of 26 items across four subscales: Sensitivity (e.g., warmth, attentiveness), Harshness (e.g., critical of children, punitiveness), Permissiveness, and Detachment (e.g., low levels of interaction with children). Each item is rated on a scale from 1 (Never true) to 4 (Often true).

The *ECERS-Extension*, or *ECERS-E* (Sylva, Siraj-Blatchford, & Taggart, 2003) contains 15 items. These items are grouped into four subscales: Literacy, Mathematics, Science and Environment, and Diversity. The *ECERS-E* was created as a more cognitively-oriented supplement to the *ECERS-R* to correspond to a newly adopted preschool curriculum in England.

It also focuses more heavily on cultural and intellectual diversity than the *ECERS-R* (Sylva, et al., 2006).

#### Factor Structure of the ECERS-R

Using factor analytic techniques, several studies have found that there is an underlying factor structure in the ECERS-R, beyond the subscale level (e.g., (R. M. Clifford & H.-G. Rossbach, in press; Early, et al., 2005; Sakai, Whitebook, Wishard, & Howes, 2003) (CITE NCEDL, CQO). Most studies examining the factor structure have identified between two and five factors.

Most commonly, two factors have emerged: Teaching and Interaction and Provisions for Learning. The Teaching and Interaction factor relates to teacher behaviors that support children's development, including supporting language-reasoning experiences, supervision and scheduling of activities, providing varied social experiences for children, and creating a warm, friendly environment for interactions. The second factor, Provisions for Learning relates to the space and materials available to children in the environment, including room arrangement, furniture, materials for fine motor play, art materials, blocks, sensory materials, and dramatic play materials. Items associated with each factor are displayed below:

**Table 1: ECERS-R Factors: Teaching and Interaction**

<b>Items</b>	Sakai et al. (2003) (n=68) (*subscale reliability correlations)	Cassidy, Hestenes, Hestenes, & Mims (2003) (n=1313) (factor loadings)	Perlman, Zellman, & Le (2004) (n=326) (factor loadings)
<b>Loadings</b>			
<b>Factor 1</b>			
Greeting/departing	.54		.45
Use language	.23	.47	.67
Informal use of language	.70	.63	.82
Group time		.48	
Staff-child interactions	.58	.72	.80
Discipline	.74	.78	.74
Interactions among children	.55	.72	.72
General supervision	.60	.54	.72
Supervision of gross motor	.44		.59
Encouraging communication	.59		.57
Health practices			.34
<hr/>			
Internal consistency Factor 1	.84	.81	



**Table 2: ECERS-R Factors: Provisions for Learning**

<b>Items</b>	<b>Sakai et al. (2003) (n=68)</b>	<b>Cassidy, Hestenes, Hestenes, &amp; Mims (2003) (n=1313)</b>	<b>Perlman, Zellman, &amp; Le (2004) (n=326)</b>
<b>Loadings</b>			
<b>Factor 2</b>			
Furniture for relaxation		.65	.60
Room arrangement	.56		
Fine motor	.43	.77	.70
Art	.60	.74	.76
Blocks	.64	.60	.67
Sand/water	.69	.59	
Dramatic play	.69		.80
Space for privacy	.44	.58	.61
Math/Number		.74	.54
Nature/Science		.73	.69
Books and pictures		.61	.48
Music/movement			.63
Free play			.55
Sand/water			.54
Schedule			.51
<hr/>			
Internal consistency			
Factor 2	.81	.87	

In addition to the overall score being stable, the mean scores for the two factors appear to remain stable over time (R. Clifford & H.-G. Rossbach, in press; Clifford, 2005). Teaching and Interactions had mean scores of 4.43 (S.D. 1.29) and 4.44 (S.D. 1.22) in the fall 2001 and spring 2002 respectively with a fall-spring correlation of .60. Provisions for Learning had scores of 3.79 (S.D. 0.96) and 3.79 (S.D. 0.88), with a correlation of .72 over this time period. Thus, these results suggest that both the factors and the full scale can maintain a high level of stability over time (R. Clifford & H.-G. Rossbach, in press; Clifford, 2005).

A confirmatory analysis was conducted to further examine the two factors across data gathered in the U.S., Germany, Portugal, and Spain. The analyses were planned according to a two step process. First, the larger U.S. data set from the Cost, Quality and Child Outcomes study (n=255 classrooms) was randomly split into two parts: a sample consisting of two thirds of the classrooms and a second part consisting of the other one third. For the two thirds sample, confirmatory factor analyses were conducted in order to provide for a first test of the hypothesized structure of two correlated factors (R. Clifford & H.-G. Rossbach, in press; Clifford, 2005). Data from these analyses are shown in Table 3. The findings indicate a factor structure that is stable across both the U.S. samples, as well as across the samples from the other countries.

**Table 3: Results of confirmatory factor analyses; loadings, correlations of factors, percentages of variance accounted for (AGFI) and internal consistencies**

Items	2/3 US sample (n=255)	1/3 US sample (n=128)	Germany (n=103)	Portugal (n=84)	Spain (n=79)
<b>Factor 1</b>					
1. Greet./depart.	.57	.63	.39	.47	.42
11. Under. Language	.85	.72	.62	.48	.75
12. Use language	.85	.84	.73	.67	.81
13. Reasoning	.83	.72	.66	.65	.71
14. Infor. Use langu.	.84	.85	.82	.56	.82
16. Superv. fine mot.	.71	.72	.68	.68	.74
26. Sched. Creative	.77	.78	.72	.55	.80
27. Superv. Creative	.73	.79	.74	.62	.75
29. Free play	.83	.83	.76	.65	.71
30. Group time	.72	.64	.74	.50	.83
32. Tone	.80	.84	.75	.62	.68
<b>Factor 2</b>					
7. Fur. Learning	.82	.77	.67	.74	.79
8. Fur. relaxation	.64	.51	.64	.28	.65
9. Room arrange	.74	.78	.85	.44	.77
15. Fine motor	.71	.78	.65	.64	.68
21. Art	.81	.77	.70	.58	.85
23. Blocks	.69	.74	.69	.59	.65
24. Sand/water	.65	.62	.41	.54	.62
25. Dramatic play	.59	.52	.69	.40	.65
28. Space to be alone	.70	.56	.75	.60	.51
Correlation of factors	.80	.73	.76	.66	.83
Percentage of variance accounted for (AGFI)	.84	.70	.68	.70	.64
Internal consistency (alpha)					
Factor 1	.94	.94	.91	.85	.92
Factor 2	.89	.88	.88	.77	.89

### Cautions in Using and Interpreting Scores from the ECERS-R

There are several cautions about interpreting the scores from the Environment Rating Scales, particularly when comparing results across studies. Most of these cautions involve the omission of some items and the calculation of the total and subscale scores.

First, the intention of the authors of the Environment Rating Scales was for the total score to be calculated by summing the scores from each item, and then finding a mean item score. Some studies instead have used a mean of the subscale scores to calculate the total score. The use of subscales rather than items shifts the weighting of the items in the overall score calculation. Rather than each item carrying the same weight, those items that are part of subscales with fewer numbers of items (e.g., items in the Program Structure subscale) carry as much weight as items that are part of larger subscales (e.g., items in the Activities subscale). It was the intention of the authors that each item is weighted equally in the total score calculation. – cite examples here

Second, it has become a common practice to omit some items altogether. While it may not always be possible to include all items, this omission should be explicitly stated as it may affect the total score. For example, some studies (e.g., (Clifford, et al., 2005; Early, et al., 2005) have omitted items from the Parents and Staff subscale. These items tend to receive relatively high scores compared to other items in the scale, and the omission of the items from this subscale can make the total scores appear to be lower than they would have been had all items been included. Occasionally, researchers will omit a single item from the scale due to concerns over measuring the item. For example, the NCEDL study omitted the Toileting item from the overall score. This decision was made due to the use of a more strict rule for hygienic conditions than is

typically applied using the ECERS-R (Clifford, et al., 2005). While it may be necessary to omit single items from the computation of the total score, any incidence of this omission should be noted as it may impact the ability of researchers to compare scores across studies.

A similar issue involves the use of only one or two subscales to represent the quality of an environment. The Environment Rating Scales were designed as a measure of the global quality of an environment for children. The authors do not recommend using individual subscales as indicators of quality because there have not been systematic studies to confirm whether the individual subscales in isolation are indicative of quality.

Additionally, care should be taken to examine the rates of attrition for samples when comparing quality scores across studies. Differential rates of attrition can skew the stability of quality over time.

## References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Psychological Association, American Educational Research Association, & National Council on Measurement in Education. (1954). *Technical recommendations for psychological tests and diagnostic techniques*. Washington, DC: The Association.
- Angoff, W. H. (1988). Validity: An evolving concept. In H. Wainer & H. Braun (Eds.), *Test Validity* (pp. 19-32). Hillsdale, NJ: Lawrence Erlbaum.
- Arnett, J. (1989). Caregivers in day care centres: Does training matter? *Applied Developmental Psychology, 10*, 541-552.
- Bailey, D. B. (2004). Tests and test development *Assessing infants and preschoolers with special needs* (3rd ed., pp. 22-44). Upper Saddle River, NJ: Pearson.
- Bryant, D., Maxwell, K., Taylor, K., Poe, M., Peisner-Feinberg, E., & Bernier, K. (2003). Smart Start and Preschool Child Care Quality in NC: Change Over Time and Relation to Children's Readiness. In K. Maxwell (Ed.). Chapel Hill, NC: FPG Child Development Institute.
- Burchinal, M., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., et al. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher-child interactions and instruction.
- Cassidy, D. J., Hestenes, L. L., Hegde, A., Hestenes, S., & Mims, S. (2005). Measurement of quality in preschool child care classrooms: An exploratory and confirmatory factor analysis of the Early Childhood Environment Rating Scale-Revised. *Early Childhood Research Quarterly, 20*, 345-360.
- Clifford, R., & Rossbach, H.-G. (in press). Structure and stability of the Early Childhood Environment Rating Scale. In H.-G. Rossbach (Ed.), *A world of improvement: Promoting quality early childhood programs for all children*.
- Clifford, R. M. (2005). Questions of quality *Center for early childhood development and education*. St. Patrick's College, Dublin, Ireland.
- Clifford, R. M., Barbarin, O., Chang, F., Early, D., Bryant, D., Howes, C., et al. (2005). What is pre-kindergarten? Characteristics of public pre-kindergarten programs. *Applied Developmental Science, 9*(3), 126-143.
- Clifford, R. M., & Rossbach, H.-G. (2005). Structure and stability of the Early Childhood Environment Rating Scale. In H. Schonfeld, S. O'Brien & T. Walsh (Eds.), Questions of Quality (pp12-21). Dublin, Ireland: The Centre for Early Childhood Development and Education, St. Patrick's College.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth Group/Thomson Learning.
- Early, D., Barbarin, O., Bryant, D., Burchinal, M., Chang, F., Clifford, R., et al. (2005). Pre-kindergarten in eleven states: NCEDE's multi-state study of pre-kindergarten and study of State-Wide Early Education Programs (SWEEP): Preliminary descriptive report. In O. Barbarin (Ed.). Chapel Hill, NC: FPG Child Development Institute.

- Harms, T., & Clifford, R. (1980). *Early Childhood Environment Rating Scale (ECERS)*. New York: Teachers College Press.
- Harms, T., & Clifford, R. (1983). Assessing preschool environments with the Early Childhood Environment Rating Scale. *Studies in Educational Evaluation*, 8, 261-269.
- Harms, T., Clifford, R., & Cryer, D. (1998). *Early Childhood Environment Rating Scale (Rev. ed.)*. New York: Teachers College Press.
- Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238-247.
- Hightower, A. D., Work, W. C., Cowen, E. L., Lotyczewski, B. S., Spinell, A. P., Guare, J. C., et al. (1986). Teacher-Child Rating Scale 2.1. *School Psychology Review*, 15, 393-409.
- Kubiszyn, T., & Borich, G. (2003). *Educational testing and measurement: Classroom application and practice* (7th ed.). Hoboken, NJ: John Wiley & Sons, Inc.
- La Paro, K. M., Pianta, R. C., & Stuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the prekindergarten year. *Elementary School Journal*, 104, 409-426.
- Love, J. M., Constantine, J., Paulsell, D., Boller, K., Ross, C., Raikes, H., et al. (2004). The role of Early Head Start programs in addressing the child care needs of low-income families with infants and toddlers: Influences on child care use and quality. In J. Constantine (Ed.): *Mathematica*.
- Mashburn, A. J., Pianta, R., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., et al. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79, 732-749.
- Melhuish, E., & Petrogiannis, K. (1996). Aspects of quality in Greek day care centres. *European Journal of Psychology of Education*, 11, 177-191.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
- Montes, G., Hightower, A. D., Brugger, L., & Moustafa, E. (2005). Quality child care and socio-emotional risk factors: No evidence of diminishing returns for urban children. *Early Childhood Research Quarterly*, 20, 361-372.
- Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., et al. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, 72, 1534-1553.
- Phillipsen, L. C., Burchinal, M. R., Howes, C., & Cryer, D. (1997). The prediction of process quality from structural features of child care. *Early Childhood Research Quarterly*, 12, 281-303.
- Pianta, R., Howes, C., Burchinal, M. R., Bryant, D., Clifford, R., Early, D., et al. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science*, 9(3), 144-159.
- Pianta, R., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS)*. Baltimore, MD: Brookes, Inc.

- Ritchie, S., Howes, C., Kraft-Sayre, M., & Weister, B. (2001). *Emerging academics snapshot*. University of California at Los Angeles. Los Angeles.
- Sakai, L. M., Whitebook, M., Wishard, A., & Howes, C. (2003). Evaluating the Early Childhood Environment Rating Scale (ECERS): Assessing differences between the first and revised edition. *Early Childhood Research Quarterly, 18*, 427-445.
- Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., & Elliot, K. (2003a). Measuring the impact of pre-school on children's cognitive progress over the pre-school period. In K. Sylva (Ed.), *The effective provision of pre-school education (EPPE) project*. London: Institute of Education.
- Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., & Elliot, K. (2003b). Measuring the impact of pre-school on children's social/behavioural development over the pre-school period. In K. Sylva (Ed.), *The effective provision of pre-school education (EPPE) project*. London: Institute of Education.
- Smith, M. W., Dickinson, D. K., Sangeorge, A., & Anastasopoulos, L. (2002). *ELLCO: User's guide to the early language and literacy classroom observation toolkit (research edition)*. Baltimore: Paul H. Brookes.
- Sylva, K., Siraj-Blatchford, I., & Taggart, B. (2003). *Assessing quality in the early years: Early Childhood Environment Rating Scale-Extension (ECERS-E): Four curricular subscales*. Stoke-on Trent: Trentham Books.
- Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., et al. (2006). Capturing quality in early childhood through environmental rating scales. *Early Childhood Research Quarterly, 21*, 76-92.
- Tenopyr, M. L. (1977). Content-construct confusion. *Personnel Psychology, 30*, 47-54.
- Tietze, W., Schuster, K.-M., Grenner, K., & Rossbach, H.-G. (2001). Kindergarten-Skala: revidierte Fassung (KES-R) [German version of the Early Childhood Environment Rating Scale Revised Edition von Thelma Harms/Richard M. Clifford/Debby Cryer}. Neiwied: Luchterhand.
- Turner, R.C., Mulvenon, S.W., Thomas, S.P., & Balkin, R.S. (n.d.) Computing indices of item congruence for test development validity assessments. SAS Statistics and Data Analysis document Paper 255-27.
- Whitebook, M. (1989). Who cares? Child care teachers and the quality of care in America. Final report, National Child Care Staffing Study. Berkeley, CA: Child Care Employee Project.
- Woodcock, R. W., & Johnson, M. B. (1990). *Woodcock-Johnson Psycho-Educational Battery-Revised*. Allen, TX: DLM Teaching Resources.
- Zill, N., & Resnick, G. (1998). *Family and Child Experiences Survey (FACES)*. Westat for Head Start National Study.



**TABLE 4. Correlations between ECERS-R subscales and other instruments**

	ECERS-R	Space and Furnishings	Personal Care Routines	Language-Reasoning	Activities	Interaction	Program Structure	Parents and Staff	Interactions Factor	Provisions Factor	Total Score
CLASS	CLASS Emotional Climate *	.31	.38	.47	.30	.58	.34	--	.588 a	.34 a	.52
	CLASS Instructional Support *	.24	.20	.52	.28	.44	.26	--	.41 a	.18 a	.40
	Positive Climate *	.27	.37	.42	.28	.55	.28	--			--
	Negative Climate *	-.21	-.27	-.37	-.26	-.52	-.27	--			--
	Sensitivity *	.23	.40	.48	.17	.50	.21	--			--
	Overcontrol *	.23	ns	.20	.26	.18	.29	--			--
	Behavior Management *	.29	.38	.42	.24	.56	.28	--			--
	Productivity *	.23	.23	.47	.24	.46	.28	--			--
	Concept Development *	.18	ns	.44	.23	.35	.21	--			--
	Learning Formats *	.18	.19	.51	.30	.44	.27	--			--
	Quality of Feedback *	.18	.18	.39	.21	.29	.15	--			--
	CIS b	Positive	.31	.29	.64	.42	.68	.44	.42		

	Relationship										
	Punitiveness	-.15	Ns	-.21	Ns	-.36	-.23	-.20			
	Permissive	-.15	Ns	-.47	-.25	-.42	-.20	-.19			
	Detachment	-.34	-.20	-.48	-.26	-.68	-.42	-.30			
CIS c	Total Score										.69
ELLCO c	Literacy activities										ns
	Literacy environmental checklist										.44
	Classroom observation total										.41
ECERS-E d											.78

\* La Paro et al., 2004

a (Pianta, et al., 2005)

b Sammons, et al., 2003)

c Peisner-Feinberg, personal communication

d Sylva, et al., 2003

