



# **Selected Observational Measures for Assessing the Quality of Early Childhood Classrooms: An Annotated Bibliography**

**May 2009**

**Martha Zaslow, Nicole Forry, Debra Weinstein,  
Mirjam Nuenning, Meagan McSwiggan, and Malissa Durham**

**Child Trends**

This document was prepared under a contract with the U.S. Department of Education's Institute of Education Sciences (IES), Contract ED-06-CO-0021, by Regional Educational Laboratory Appalachia, administered by CNA. The content of the response does not necessarily reflect the views or policies of IES or the U.S. Department of Education, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

---

## **Table of Contents**

<i>Overview</i> .....	2
Summary.....	3
<i>Terminology: Brief Descriptions</i> .....	5
Sources.....	6
<i>Descriptions of the Selected Measures</i> .....	7
Early Childhood Environment Rating Scale–Revised (ECERS-R).....	8
Classroom Assessment Scoring System (CLASS).....	15
Early Childhood Environment Rating Scale–Extension (ECERS-E).....	22
Early Language and Literacy Classroom Observation Toolkit (ELLCO).....	28
<i>Annotated Bibliography</i> .....	34
ECERS-R.....	34
CLASS.....	45
ECERS-E.....	51
ELLCO.....	53
Emerging Work.....	58

---

## Overview

This document is an annotated bibliography summarizing recent research articles that focus on four measures for observing the quality of early childhood classrooms: the Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms, Clifford, and Cryer, 1998); the Classroom Assessment Scoring System (CLASS; Hamre, Mashburn, Pianta, Locasale-Crouch, and LaParo, 2006); the Early Childhood Environment Rating Scale-Extension (ECERS-E; Sylva, Siraj-Blatchford, and Taggart, 2003); and the Early Language and Literacy Classroom Observation Toolkit (ELLCO; Smith, Dickinson, Sangeorge, and Anastasopoulos, 2002). The first two measures described provide broad descriptions of classroom quality, while the second two concentrate more specifically on the quality of instruction within the classroom. The summarized research articles regarding each of the specific measures of quality are all published journal articles that provide information on the reliability and validity of the measures. In a final section we briefly summarize emerging work, still under review prior to publication that looks across observational measures of quality in early childhood classrooms and offers comparative information on their functioning.

To provide context for the summaries of recent research articles, we begin this document with a brief description of terminology used, such as definitions for reliability and validity and overviews of approaches to measuring these. We also include overviews of the four selected measures, including the purpose for which each was developed, the population for whom each is appropriate, how each measure addresses diversity within the classroom, what constructs it covers, what is required for administration of each measure, and the evidence presented by the authors of each measure regarding its reliability and validity. These descriptions are taken directly from a publicly available document prepared by Child Trends researchers Tamara Halle, Jessica Vick and colleagues (2007): *Quality in Early Childhood Care and Education Settings: A Compendium of Measures*, Washington, DC. Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

(<http://www.researchconnections.org/location/13403>). Before turning to the glossary, description of each of the four selected measures, and the annotated bibliography itself, we present a summary of key findings from the articles profiled in the annotated bibliography.

## Summary

In a recent review of the bases for selecting among measures of quality for early childhood classrooms, Bryant (under review) notes the following as factors that should be considered: the purposes of the quality assessment; the domains of development considered to be important and the relative emphasis being placed on predicting academic school readiness; evidence of reliability and validity for the measure; evidence that the measure can detect changes in quality with program improvement interventions; evidence of appropriateness of the measure for the populations included in the programs for which quality is being measured; and logistic considerations, such as cost of obtaining the measure, cost of both training observers and maintaining, reliability.

The ECERS-R, CLASS and ELLCO are among the most frequently used observational measures of classroom quality. They differ in the extent to which they provide a broad portrayal of quality (ECERS-R) or focus more specifically on instructional practices (CLASS; ELLCO). A recent extension of the ECERS-R, the ECERS-E, adds ratings of instructional quality in early literacy, science, and mathematics to the global measure of quality.

There is evidence of reliability and validity for each of the observational measures of quality we have included (Bryant, Clifford, and Peisner, 1991; Burchinal and Cryer, 2003; Burchinal, Howes, Pianta, Bryant, Early, Clifford, and Barbarin, 2008; Cassidy, Hestenes, Hedge, Hestenes, and Mims, 2005; Dickinson and Caswell, 2007; Gettinger and Stoiber, 2007; Hindman and Wasik, 2008; Jackson, Larzelere, St. Clair, Corr, Fichter, and Egertson, 2006; LaParo, Pianta, and Stuhlman, 2004; Perlman, Zellman, and Le, 2004; Sakai, Whitebook, Wishard, and Howes, 2003; Sylva, Siraj-Blatchford, Taggart, Sammons, Melhuish, Elliot, and Totsika, 2006; Sylva, Taggart, Siraj-Blatchford, Totsika, Ereky-Stevens, Gilden, and Bell, 2007). Evidence of validity includes at least some evidence for each measure's prediction of child outcomes (Burchinal and Cryer, 2003; Bryant, under review; Burchinal, Howes, Pianta, Early, Clifford, and Barbarin, 2008; Burchinal, Kainz, and Cai, under review; Howes, Burchinal, Pianta, Bryant, Early, Clifford, and Barbarin, 2008; LaParo, Pianta, and Stuhlman, 2004; Mashburn, Pianta, Hamre, Downer, Barbarin, Bryant, Burchinal, Early, and Howes, 2008; Sylva, Siraj-Blatchford, Taggart, Sammons, Melhuish, Elliot, and Totsika, 2006; Jackson, Larzelere, St. Clair, Corr, Fichter, and Egertson, 2006). However, the measures vary in terms of which child outcomes they predict and the strength of these predictive relationships. There is evidence that measures with a focus on instructional practice do a somewhat better job of predicting academic achievement outcomes than global measures of quality (Burchinal, Kainz, and Cai, under

review; Howes, Burchinal, Pianta, Bryant, Early, Clifford, and Barbarin, 2008; Mashburn, Pianta, Hamre, Downer, Barbarin, Bryant, Burchinal, Early, and Howes, 2008; Sylva, Siraj-Blatchford, Taggart, Sammons, Melhuish, Elliot, and Totsika, 2006).

A recent review on the measurement of quality in early childhood classrooms notes that using one of the measures that emphasizes instructional quality alone eliminates the focus on health and safety and, depending upon the measure, can also diminish the focus on adequacy of facilities and equipment (Bryant, under review). Some evaluations of quality improvement programs or policy initiatives use multiple measures of quality to capture all aspects included in global quality measures, as well as a more detailed consideration of instructional quality through combining the ECERS-R with the CLASS, ELLCO, or ECERS-E. As one example, the evaluation of the Missouri Quality Rating system (Thornburg, Mauzy, Mayfield, Scott, Sparks, Mumford, Foulkes, and Furger, under review) includes both the ECERS-R and the ECERS-E.

---

## **Terminology: Brief Descriptions**

**Developmentally appropriate:** “Developmentally appropriate practice” is practice that “is informed by what is known about child development and learning, what is known about each child as an individual, and what is known about the social and cultural contexts in which children live (adapted from National Association for the Education of Young Children, 1996, 2008).” (Snow and Van Hemel, 2008, p. 425).

**Reliability:** “The consistency of measurements, gauged by any of several methods, including when the testing procedure is repeated on a population of individuals or groups (test-retest reliability), or is administered by different raters (inter-rater reliability). There is no single, preferred approach to quantification of reliability (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999)” (Snow and Van Hemel, 2008, p. 427).

*Inter-rater Reliability:* The degree of consistency among raters. A common way of assessing inter-rater reliability, which takes into consideration chance agreement, is the kappa statistic (Neuman, 1997).

*Internal Consistency:* The degree of consistency in results across items within a test (Neuman, 1997).

*Stability Across Time:* The degree to which a measure delivers the same answer across time periods. One common way of assessing this is the test-retest method (Neuman, 1997).

**School readiness test:** “A testing instrument designed to measure skills believed to be related to school learning tasks and to be predictive of school success (Council of Chief State School Officers, 2008)” (Snow and Van Hemel, 2008, p. 426).

**Validity:** “The extent to which an instrument measures what it purports to measure; the extent to which an assessment’s results support meaningful inferences for certain intended purposes.” (Snow and Van Hemel, 2008, p. 427).

*Criterion Validity:* “Criterion validity uses some standard (or criterion) that is known to indicate a construct accurately. The validity of a measure is verified by comparing it with another measure of the same construct in which a researcher has confidence” (Neuman, 1997, p. 144). There are two types of criterion validity: concurrent validity and predictive validity.

Concurrent Validity: “The extent to which an indicator is associated with a preexisting indicator already judged to be valid (i.e. has face validity)” (Neuman, 1997, p. 144).

Predictive Validity: The extent to which an indicator “predicts future events that are logically related to a construct” (Neuman, 1997, p. 144).

*Construct Validity*: The degree to which measures with multiple indicators operate consistently. Construct validity can be evaluated in terms of convergent validity (degree to which “multiple indicators are associated with indicators measuring the same construct”) and discriminant validity (degree to which “indicators are negatively associated with opposing constructs”) (Neuman, 1997, p. 144).

*Content Validity*: An assessment of the degree to which the “full content of a definition is represented in a measure” (Neuman, 1997, p. 142).

## Sources

Neuman, W. L. (1997). *Social Research Methods: Qualitative and Quantitative Approaches*, 3rd Edition. Boston: Allyn and Bacon.

Snow, C. E. and Van Hemel, S. B. (2008). *Early childhood assessment: Why, what, and how*. Washington, DC: National Academies Press.

---

## ***Descriptions of the Selected Measures***

This section provides a detailed overview of the four measures included in this review. The measures are presented in the following order: Early Childhood Environment Rating Scale-Revised (ECERS-R), Classroom Assessment Scoring System (CLASS), Early Childhood Environment Rating Scale- Extension (ECERS-E), and Early Language and Literacy Classroom Observation Toolkit (ELLCO). This order reflects the specificity of the measures, beginning with a broad classroom observation covering multiple domains (ECERS-R) to a specific language and literacy curriculum and materials observation (ELLCO).

The information provided in this section was taken directly from Halle, T., and Vick, J. E. (2007). *Quality in Early Childhood Care and Education Settings: A Compendium of Measures*. Washington, DC: Prepared by Child Trends for the Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. This publication is available online at: [http://www.childtrends.org/Files/Child\\_Trends-2007\\_12\\_10\\_FR\\_CompleteCompendium.pdf](http://www.childtrends.org/Files/Child_Trends-2007_12_10_FR_CompleteCompendium.pdf) or <http://www.researchconnections.org/location/13403>. The full compendium includes profiles for additional observational measures than those extracted here, and an update of the compendium is being prepared to include newly developed measures of quality. The costs noted in this extract from the compendium are based on information collected in 2007. The updated compendium will update information on costs of manuals and training materials.



---

## Early Childhood Environment Rating Scale–Revised (ECERS-R)

### Background Information

#### Author/Source

*Source:* Harms, T., Clifford, R. M., and Cryer, D. (1998). *Early Childhood Environment Rating Scale – Revised Edition*. New York, NY: Teachers College Press.  
Harms, T., Clifford, R. M., and Cryer, D. (2005). *Early Childhood Environment Rating Scale – Revised Edition*. New York, NY: Teachers College Press. (Updated with additional notes and an expanded score sheet).

*Publisher:* Teachers College Press  
1234 Amsterdam Avenue  
New York, New York 10027

#### Purpose of Measure

*As described by the authors:*

The Early Childhood Environment Rating Scale (ECERS-R) measures global quality in center-based early childhood programs. The ECERS-R can be used as a tool “to see how well a program is meeting children’s needs – to see whether children receive the protection, learning opportunities, and positive relationships they need for successful development” (Cryer, Harms, and Riley, 2003, p. x). It can be used by researchers, practitioners, program monitors, and early childhood professionals who provide technical assistance to programs.

The ECERS-R is a revision of the ECERS originally published in 1980. “The ECERS-R retains the original scale’s broad definition of environment, including those spatial, programmatic, and interpersonal features that directly affect the children and adults in an early childhood setting” (Harms, Clifford, and Cryer, 1998, p. 1).

#### Population Measure Developed With

Information not available in materials reviewed.

#### Age Range/Setting Intended For

The ECERS-R is designed to be used with one room or one group at a time, for children two-and-one-half through five years of age in center-based programs.

#### Ways in which Measure Addresses Diversity

- Indoor Space (item # 1) assesses whether the space is accessible to children and adults with disabilities.

- Furniture for Routine Care, Play, and Learning (item #2) assesses whether children with disabilities have adaptive furniture that facilitates their inclusion in classroom activities.
- Room Arrangement for Play (item # 4) assesses whether play spaces are accessible to children with disabilities.
- Space for Gross Motor Play (item #7) assesses whether the gross motor space is accessible for children in the group.
- Gross Motor Equipment (item # 8) assesses whether adaptations are made or special equipment is provided for children with disabilities.
- Meals/Snacks (item #10) assesses whether children with disabilities are included at the table with their peers and whether dietary restrictions of families are followed.
- Toileting and Diapering (item # 12) assesses whether provisions are convenient and accessible for children.
- Books and Pictures (item #15) assesses whether there are a variety of books in the classroom and whether they reflect different cultures and abilities.
- Music/Movement (item #21) assesses whether music materials are adapted for children with disabilities and whether music from different cultures and in different languages is represented.
- Dramatic Play (item # 24) assesses whether props, such as dolls and dress-up clothes, are provided to represent diversity of cultures and abilities.
- Promoting Acceptance of Diversity (item # 28) assesses whether the materials and activities represent and positively portray different races, cultures, ages, genders, and abilities.
- Provisions for children with disabilities (item # 37) assesses whether modifications are made in the environment to allow children with disabilities to participate fully and be integrated into the group. The item also assesses whether teachers interact with parents and specialists to plan for meeting the child's needs.

### **Key Constructs & Scoring of Measure**

The scale consists of 43 items categorized into seven subscales. Items are scored on a seven-point scale from 1 to 7. Numbered indicators outlining the specific requirements for the item are provided at score points 1 (inadequate), 3 (minimal), 5 (good), and 7 (excellent). The observer begins at level 1 and scores each indicator “yes,” “no,” or “NA.” The final score is determined by the number of indicators that have been “passed.” All indicators must be passed at each level to score at or above that level.

Thus, to score a 7 on an item, all indicators must be passed including all of those included under Level 7.

- Space and Furnishings (8 items)
  - Indoor space
  - Furniture for routine care, play, and learning
  - Furnishings for relaxation and comfort
  - Room arrangement for play
  - Space for privacy
  - Child-related display
  - Space for gross motor play
  - Gross motor equipment
- Personal Care Routines (6 items)
  - Greeting/departing
  - Meals/snacks
  - Nap/rest
  - Toileting/diapering
  - Health practices
  - Safety practices
- Language-Reasoning (4 items)
  - Books and pictures
  - Encouraging children to communicate
  - Using language to develop reasoning skills
  - Informal use of language
- Activities (10 items)
  - Fine Motor
  - Art
  - Music/movement
  - Blocks
  - Sand/water
  - Dramatic play
  - Nature/science
  - Math/number
  - Use of TV, video, and/or computers
  - Promoting acceptance of diversity

- Interaction (5 items)
  - Supervision of gross motor activities
  - General supervision of children (other than gross motor)
  - Discipline
  - Staff-child interactions
  - Interactions among children
- Program Structure (4 items)
  - Schedule
  - Free play
  - Group time
  - Provisions for children with disabilities
- Parents and Staff (6 items)
  - Provision for parents
  - Provision for personal needs of staff
  - Provision for professional needs of staff
  - Staff interaction and cooperation
  - Supervision and evaluation of staff
  - Opportunities for professional growth.

### **Comments**

The ECERS-R contains Notes for Clarification on each item that define the terms used in the item and clarify specific scoring requirements for the indicators that comprise the item. There are also Additional Notes for the ECERS-R that provide more detailed information to be considered in scoring and address scoring questions that the authors have answered since publication of the scale. The Additional Notes can be found at the following website: <http://www.fpg.unc.edu/~ecers/> or in the updated 2005 ECERS-R book.

### **Administration of Measure**

#### **Who Administers Measure/Training Required**

*Test Administration:* The ECERS-R book provides questions for each item that can guide the interview. The authors also provide specific instructions for administering the scale and for conducting the observation in a way that minimizes the impact of the observer on the classroom environment. Because of the large number of indicators that need to be scored, the observer should have the ECERS-R book with her/him while in the classroom and should complete scoring before leaving the facility.

*Training Required:* The authors recommend that observers “participate in a training sequence led by an experienced ECERS-R trainer before using the scale formally. The training sequence for observers who will use the scale for monitoring, evaluation, or research should include at least two practice classroom observations with a small group of observers, followed by inter-rater reliability comparison” (Harms, Clifford, and Cryer, 1998, p. 5). Five-day and three-day trainings are offered by the authors of the scale at the University of North Carolina, Chapel Hill. Observers can purchase additional resources including a video training package (available from Teachers College Press) or the *All About the ECERS-R* book (Cryer, Harms, and Clifford, 2003) that offers detailed information and photos to assist the observer in learning the scale or interpreting and scoring what s/he has seen in a classroom. The authors note the use of *All About the ECERS-R* will assist groups of ECERS-R observers in developing reliability and being more consistent with the ECERS-R authors.

### **Setting**

Observations are made in classrooms within center-based settings, including child care centers, pre-schools, nursery schools, and pre-kindergarten programs.

### **Time Needed and Cost**

*Time:* The ECERS-R should be used by a trained observer at a time when children are awake and active. The observation should include “both play/learning times and routines, such as a meal, toileting, and preparation for nap” (Cryer, Harms, and Riley, 2003, p. xiv). The authors recommend that at least two-and-one-half to three hours be spent observing in the classroom and note that spending more than three hours observing is preferable. An additional 20-30 minutes is needed to ask the teacher questions to help score indicators that were not observed.

*Cost:* All materials are available through Teachers College Press Manuals

(ECERS-R, 2005) \$17.95

(ECERS-R, 1998) \$14.95

Video Training Packages

1999, VHS \$59.00

2006, DVD \$59.00

Training Workbook

1999 \$4.00

## Functioning of Measure

### Reliability Information

#### *Inter-rater Reliability*

“Overall the ECERS-R is reliable at the indicator and the item level, and at the level of the total score. The percentage of agreement across the full 470 indicators in the scale is 86.1%, with no item having an indicator agreement level below 70%. At the item level, the proportion of agreement was 48% for exact agreement and 71% for agreement within one point. For the entire scale, the correlations between the two observers were .92 product moment correlation (Pearson) and .87 rank order (Spearman). The interclass correlation was .92” (Harms, Clifford, and Cryer, 1998, p. 2).

#### *Internal Consistency*

The authors “also examined the internal consistency of the scale at the subscale and total score levels. Subscale internal consistencies range from .71 to .88 with a total scale internal consistency of .92” (Harms, Clifford, and Cryer, 1998, p. 2).

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 Information

#### *Predictive Validity*

The authors note that, since the original ECERS had demonstrated that “quality as measured by the ECERS has good predictive validity (i.e., Peisner-Feinberg and Burchinal, 1997; Whitebrook, Howes, and Phillips, 1990), the revised version would be expected to maintain that form of validity” (Harms, Clifford, and Cryer, 1998, p. 2).

#### *Content Validity*

When the scale was revised, the authors conducted focus groups with experts in the field who made suggestions for the revision based on how the ECERS had worked in inclusive

and culturally diverse settings. The authors also gathered feedback and suggestions from researchers and other ECERS users that informed the content in the ECERS-R.

## References

Cryer, D., Harms, T., and Riley, C. (2003). All about the ECERS-R: A detailed guide in words and pictures to be used with the ECERS-R. PACT House Publishing.

Peisner-Feinberg, E., and Burchinal, M. (1997). Relations between preschool children's child care experiences and concurrent development: The Cost, Quality and Outcomes Study. *Merrill-Palmer Quarterly*, 43, 451-477.

Whitebook, M., Howes, C., and Phillips, D. (1990). Who cares? Child care teachers and the quality of care in America. Final report of the National Child Care Staffing Study. Oakland, CA: Child Care Employee Project.

---

# Classroom Assessment Scoring System (CLASS)

## Background Information

### Author/Source

*Source:* Hamre, B. K., Mashburn, A. J., Pianta, R. C., Locasale-Crouch, J., and La Paro, K. M. (2006). *Classroom Assessment Scoring System Technical Appendix*. Pianta, R. C., La Paro, K. M., and Hamre, B. K. (2007). Classroom Assessment Scoring System. Brookes Publishing

*Publisher:* Paul H. Brookes Publishing Co.  
Post Office Box 10624  
Baltimore, MD 21285-0624  
Phone: 800-638-3775  
Website: [www.brookespublishing.com](http://www.brookespublishing.com)

### Purpose of Measure

*As described by the authors:*

The Classroom Assessment Scoring System (CLASS) is an observational instrument developed to assess classroom quality in pre-school through grade 3 classrooms. The CLASS dimensions are based on observed interactions among teachers and students in classrooms. The dimensions were derived from a review of constructs assessed in classroom observation instruments used in child care and elementary school research, literature on effective teaching practices, focus groups, and extensive piloting. The Observational Record of Classroom Environments (ORCE, ECRN, NICHD, 1996) served as a foundation for the development of the CLASS. The instrument may be used as a research tool, a professional development tool, and/or as a program development and evaluation tool.

### Population Measure Developed With

The technical appendix identifies six studies on which the psychometric information for the CLASS is based.

- 694 preschool classrooms in 11 states; 730 kindergartens in six states (National Center for Early Development and Learning MS and SWEEP studies)
- 164 preschool classrooms in Virginia (MyTeachingPartner Study)
- 82 grade 3 – grade 5 classrooms in New York City (4R's Study)
- 88 grade 1 – grade 5 classrooms in an urban district in the Northeast (Responsive Classroom Study)
- 33 classrooms (K-5) in a Southeastern city (Induction Study)



- Approximately 900 classrooms in each of grade 1, grade 3, and grade 5 in 10 sites nationally (NICHD Study of Early Child Care and Youth Development)

Collectively, the CLASS has been validated in over 3,000 classrooms throughout the United States.

### **Age Range/Setting Intended For**

The CLASS was developed for use in preschool through grade 3 classrooms. Currently two versions of the CLASS are available: a preschool version and a K-3 version. The CLASS approach provides a common metric and language for discussion of quality across age levels and grades. Versions of the CLASS for use in Infant/Toddler, Upper Elementary, and Secondary grades are currently in development. Data on these versions are available from the authors (contact Bridget Hamre, Ph.D. at [hamre@virginia.edu](mailto:hamre@virginia.edu)).

### **Ways in which Measure Addresses Diversity**

The CLASS has been used and validated in large national studies including a diverse range of classrooms and children (Howes et al., in press; Pianta et al., 2005).

### **Key Constructs and Scoring of Measure**

Ten dimensions of classroom quality are identified across three domains of interaction – Emotional Support, Classroom Organization, and Instructional Support. These domains of interaction are common across the preschool to grade 3 period. Each dimension is rated on a seven-point Likert-type scale. The manual describes anchor behaviors for Low (1,2), Mid (3,4,5), and High (6,7) scores for each item.

- Emotional Support
  - Positive Climate
  - Negative Climate
  - Teacher Sensitivity
  - Regard for Student Perspectives
- Classroom Organization
  - Behavior Management
  - Productivity
  - Instructional Learning Formats
- Instructional Support
  - Concept Development
  - Quality of Feedback

## – Language Modeling

### **Comments**

Previous versions of the CLASS have included the following constructs: Over-control (replaced by Regard for Student Perspectives), Literacy Development (replaced by Language Modeling and Literacy Focus), Quality of Numeracy and Math Instruction, Social Studies Instruction and Activities, Science Instruction and Activities, and Children’s Engagement (Hamre et al., 2006; La Paro and Pianta, 2003-R). Ratings should reflect the overall classroom environment as experienced by the children. That is, if there are multiple teachers in the room, all teacher behavior should be included to determine a rating. However, the CLASS can be easily adapted for use to describe the quality of a particular teacher. Observation notes are the primary source of supporting evidence for ratings.

### **Administration of Measure**

#### **Who Administers Measure/Training Required**

*Test Administration:* Trained CLASS users observe in classrooms for twenty-minute intervals and then score each CLASS dimension. The manual recommends gathering at least four of these twenty-minute intervals to assess a classroom. It is also possible to score with the CLASS based on videotaped footage. Although the manual describes a standardized protocol for observation, the procedure can be modified to meet the goals of specific projects.

*Training Required:* Training is required to assure proper use of the instrument for each of its intended uses (i.e., research, professional development, program development, and evaluation). All observers must attend training and pass a reliability test. Regular training sessions are available at the University of Virginia and the University of North Carolina – Greensboro. Personnel are also available to provide local trainings. In addition, the Train-the-Trainer Program allows representatives from universities, programs, agencies, or school districts to become certified CLASS trainers in order to train others within their organization.

#### **Setting**

Observations are made in the classroom.

#### **Time Needed and Cost**

*Time:* The authors recommend observing for a minimum of four 20-minute cycles (approximately two hours total) to obtain an accurate sampling of classroom quality data across the three CLASS domains. Total time will vary depending on the purpose of the observation.

Cost: Two-day training at the University of Virginia: \$600/person

Four-day training (Train the Trainer): \$1,000/person

Local Training: \$3,000 for up to 15 people (plus travel costs for one trainer)

PreK Manual: \$49.95

K-3 Manual: \$49.95

Pack of 10 scoring forms: \$25

## Functioning of Measure

### Reliability Information

#### *Inter-rater Reliability*

As mentioned earlier, all observers must attend training on the CLASS and take a reliability test. Observers code five 20-minute videotaped classroom sessions. The average inter-rater reliability (within one point of master codes) is reported in the Technical Appendix (p. 9) as 87 percent. Two observers both coded a total of 33 30-minute digital videotapes submitted by teachers in the MyTeachingPartner (MTP) Study. Inter-rater reliability (within one point of each other) ranged from 78.8 percent (for Behavior Management and Instructional Learning Formats) to 96.9 percent (for Productivity). Similar levels of reliability have been obtained in live observations (Hamre et al. 2006, p. 9).

#### *Internal Consistency*

Correlations among the CLASS dimensions range from .11 to .79. Correlations for the preschool sample in the MS/SWEEP Studies were generally lower than those for the grade 3 sample in the 4R's Study. Confirmatory factor analyses were performed on data from each of the studies except for the Induction Study (Hamre et al., 2006). Analyses revealed three factors representing Emotional Support, Classroom Organization, and Instructional Support. Within the MTP sample, which used the most current version of the CLASS, internal consistencies were: Emotional Support (alpha = .89); Classroom Organization (alpha = .77); and Instructional Support (alpha = .83).

#### *Stability across Time*

Stability of ratings across observation cycles was assessed in preschool and grade 3 classrooms using data from the NCEDL MS Study of preschool and the 4R's Study of grade 3 classrooms in New York City. For the grade 3 sample, correlations between the first cycle and

the total score are moderate to high, ranging from .68 for Productivity to .87 for Positive Climate. For the preschool sample, correlations between the first four cycles and the final score ranged from .84 for Productivity to .91 for Concept Development. By completing two cycles, correlations with the final score are uniformly high, with almost all correlations above .90 in both preschool and grade 3 (Hamre et al., 2006, p. 10). Correlations between observations made on two consecutive days suggest a high degree of stability, with correlations between the two days ranging from .73 for Productivity to .85 for Teacher Sensitivity.

“There were small but significant mean changes across several of the dimensions with a general trend toward lower quality scores on the second day. Given that there is no reason to expect a systematic difference in quality across two consecutive days these small changes may be due to observer bias in which scores become slightly lower over time. Again, however, although these differences are statistically significant, they are relatively small effects and correlations between the two days are high” (Hamre et al., 2006, p. 13).

CLASS scores have also been found to be relatively stable across the school year, at least in a large number of preschool classrooms. Analyses also indicate that seven-point rating scales of the classroom are highly stable and not dependent on occasion.

### **Validity Information**

#### *Criterion Validity*

The CLASS domains of Emotional Support, Classroom Organization, and Instructional Support are correlated with teacher reports of depression and adult-centered attitudes. Specifically, classrooms with lower scores across the CLASS dimensions had teachers who reported higher levels of depression, while those with lower scores on classroom organization and instructional support had teachers who reported more adult-centered attitudes.

#### *Concurrent Validity*

In comparisons of the CLASS with the Early Childhood Environment Rating Scale (ECERS-R), classrooms with higher CLASS scores were rated higher on the ECERS interactions factor (correlations range from .45 to .63). Correlations between CLASS ratings and the Furnishings and Materials factor from the ECERS were only moderate, ranging from .33 to .36 (Pianta et al., 2005).

The CLASS has also been compared to The Snapshot, a time-sampling method used to assess the percent of time spent on various activities (Pianta et al., 2005). Because the CLASS assesses the quality rather than the quantity of classroom activities, it is not surprising that there were low (but still significant) correlations between the CLASS instructional support domain and time spent in literacy and math according to The Snapshot. Children in classrooms with higher

CLASS scores spent more time in elaborated interactions with adults and significantly more time engaged.

### *Predictive Validity*

Results from the NCEDL Multi-state study provide evidence that classroom quality, as assessed by the CLASS, is associated with children's performance at the end of preschool, as well as gains in their performance across the preschool year (Howes et al., 2008). These associations were sustained, even after controlling for a variety of covariates, including maternal education, ethnicity, and gender. The most consistent and robust classroom quality dimension for predicting growth across time was the Instructional Support of the classroom as assessed by the CLASS. The CLASS Emotional Support scale was associated with growth in children's expressive and receptive language scores, as well as decreases in teacher-reported behavior problems (Howes et al., 2008).

### *Content Validity*

The CLASS dimensions are based on observed interactions among teachers and students in classrooms. The dimensions were derived from an extensive review of constructs assessed in classroom observation instruments used in child care and elementary school research, literature on effective teaching practices, focus groups, and piloting.

## **References**

- Hamre, B. K., Mashburn, A. J., Pianta, R. C., Lacasle-Crouch, J., and La Paro, K. M. (2006). *Classroom Assessment Scoring System Technical Appendix*. Classroom Assessment Scoring System (CLASS).
- Hamre, B. K., and Pianta, R. C. (2005). Can Instructional and Emotional Support in the First-Grade Classroom Make a Difference for Children at Risk of School Failure? *Child Development*, 76, 949-967.
- Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., and Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in prekindergarten programs. *Early Childhood Research Quarterly*, 23, 27-50.
- La Paro, K. M., and Pianta, R. C. (2003-R). Classroom Assessment Scoring System (CLASS) Pre-K Version. Charlottesville, VA: National Center for Early Development and Learning.
- LaParo, K. M., Pianta, R. C., and Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the pre-kindergarten year. *The Elementary School Journal*, 104(5), 409-426.
- Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., and Barbarin, O. (2005) Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child-teacher interactions? *Applied Developmental Science*, 9, 144-159.

Pianta, R. C., La Paro, K. M., and Hamre, B. K. (2007). Classroom Assessment Scoring System. Brookes Publishing.

---

## Early Childhood Environment Rating Scale–Extension (ECERS-E)

### Background Information

#### Author/Source

*Source:* 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, UK: Trentham Books.

*Publisher:* Trentham Books Limited  
Westview House, 734 London Road  
Stoke on Trent, ST4 5NP  
United Kingdom  
Phone: +44(0) 1782 745567  
E-mail: [tb@trentham-books.co.uk](mailto:tb@trentham-books.co.uk)

#### Purpose of Measure

*As described by authors:*

“The Early Childhood Environment Rating Scale – Extension (ECERS-E) was developed to supplement the ECERS-R by a team of researchers at the Institute of Education, University of London. ECERS-E reflects the English National Early Childhood Curriculum Guidance for the Foundation Stage (QCA 2000) as well as the changing notions of Developmentally Appropriate Practice.

“Four new sub-scales have been devised for the ECERS-E: Literacy, Mathematics, Science, and Diversity. Items in these sub-scales assess the quality of curricular provision, including pedagogy, in these domains aimed at fostering children’s academic development (Sammons et al., 2002)” (Sylva, Siraj-Blatchford, and Taggart, 2003, p. 7).

#### Population Measure Developed With

“The ECERS-E has been piloted extensively in a variety of settings for predictive validity (Sylva AERA, 2001). A study of 3,000 children in Britain (The Effective Provision of Pre-School Education [EPPE] Project, Institute of Education, University of London) has shown that assessments of their Early Childhood Settings made on the ECERS-E are better predictors of children’s intellectual and language progress (three-five years) than were assessments on the same settings using the ECERS-R. This validation came from a national study carried out in England to explore the relationship between the quality of the preschool measured by the Early Childhood Environment Rating Scale-Revised and the developmental progress of more than 3,000 pre-school children” (Sylva et al., 2003, pp. 7-8).

#### Age Range/Setting Intended For

The ECERS-E may be used with children three through five years of age.

## Ways in which Measure Addresses Diversity

The ECERS-E was developed in part, because the ECERS-R does little to assess diversity in the childcare setting. The ECERS-E has a “*Diversity*” subscale that assesses: caregivers’ planning for students’ individual needs, gender and equity awareness in the classroom, and race equality as reflected in materials available and caregivers’ practices. The ‘Planning for Individual Learning Needs’ item assesses how well centers plan and provide for the needs of all children in the group, whereas the ECERS-R only considers individual provision for children with identified and diagnosed special needs/disabilities.

## Key Constructs of Measure

The ECERS-E supplements the ECERS-R with four new subscales. Items are rated on a seven-point scale from (1) Inadequate to (7) Excellent. Examples are provided at scoring points 1, 3, 5, and 7 for each item. Average subscale scores can also be calculated.

- Literacy (6 items)
  - ‘Environment print’: Letters and words
  - Book and literacy areas
  - Adult reading with the children
  - Sounds in words
  - Emergent writing/mark making
  - Talking and Listening
- Mathematics (4 items)
  - Counting and the application of counting
  - Reading and writing simple numbers
  - Mathematical Activities: Shape and space (complete 3 or 4)
  - Mathematical Activities: Sorting, matching and comparing (complete 3 or 4)
- Science (5 items)
  - Natural materials
  - Areas featuring science/science resources
  - Science Activities: Science processes: non-living (complete 3, 4 or 5)
  - Science Activities: Science processes: living processes and the world around us (complete 3, 4 or 5)
  - Science Activities: Science processes: food preparation (complete 3, 4 or 5)
- Diversity (3 items)
  - Planning for individual learning needs



- Gender equity and awareness
- Race equality

## **Administration of Measure**

### **Who Administers Measure/Training Required**

*Test Administration:* The ECERS-E can be used as a self-assessment and improvement tool by well-trained observers. However, it is not generally recommended that the ECERS-E be used in isolation. It was designed as an extension to the ECERS-R to cover specific curricular areas in greater depth and not as a stand-alone tool.

#### *Training Required:*

“Before using the ECERS-E scale as either a self-assessment tool or a research instrument, it is strongly recommended that the user has some familiarity with the ECERS-R scale. The Teachers College Press have produced a range of materials to accompany these scales that have been developed for training purposes. These include video extracts and advice on making judgments. These materials can be used for both group and self-instruction. After viewing the training package, users will need to conduct several ‘trial’ observations in order to familiarize themselves with the content of the items included in the scale. This cannot be done in one observation. Using the scales demands a high degree of understanding about not only the content of the scales but about making sense of what is being observed. In many cases information to complete the scales cannot be readily observed and the user may need to question centre staff sensitively about their practices. Any user therefore needs to be familiar with the content of the scales and also to be confident in probing for additional information beyond that which is observed.

Before using the scales, users should note that it is also strongly recommended that the observer have some external validation conducted on their judgments...” (Sylva et al., 2003, p. 9).

### **Setting**

The ECERS-E may be used in early childhood classrooms, one room or one group at a time.

### **Time Needed and Cost**

*Time:* Ideally, a half-day of orientation and two guided observations are recommended for ECERS-E training. If training to use the scales to research standards, this should be followed by appropriate checks of inter-rater reliability.

For the actual observations, it is recommended that observers spend at least half a day in the classroom (and preferably longer). The authors note that observers should allow at least

15 minutes to speak with staff and children at the end of the observation to ask any additional questions.

*Cost:* The cost of training and reliability will vary depending on personnel costs. It is estimated that basic training on the ECERS-E might cost in the region of £300-400 per person (roughly \$620-\$820), and basic training on the ECERS-R and E together might cost approximately £500-600 (roughly \$1,000-\$1,200) per person. Training to research standards (i.e., with appropriate reliability checks) might cost in the region of £1,200 (roughly \$2,500) per person. The scales themselves are priced at £12.99 (roughly \$27) plus delivery costs.<sup>1</sup>

## Functioning of Measure

### Reliability Information

#### *Inter-rater reliability*

“Inter-rater reliability on the ECERS-E was calculated from data obtained from the same 25 randomly chosen centers that were also used in the reliability analysis of the ECERS-R (Sylva et al., 1999). The reliability coefficients were calculated separately for separate regions, both percentages of exact agreement between the raters and as weighted kappa coefficient. The percentages of inter-rater agreement range from 88.4 to 97.6 and the kappas range from 0.83 to 0.97...” (Sylva et al., 2003, p. 44).

#### *Internal Consistency*

“Factor analysis conducted on the ECERS-E in 141 centers (Sylva et al., 1999) indicated the presence of two factors that together account for about 50% of the total variance in the scores. The first factor is called Curriculum Areas and the second is called Diversity...”

Cronbach’s alpha was calculated for each factor and for factor 1 was high (0.84) but moderate for factor 2 (0.64). Therefore internal reliability is high only for the first factor, indicating that more factor analyses on the ECERS-E are needed. . .” (Sylva et al., 2003, pp. 44-45).

### Validity Information

#### *Construct Validity*

“In the Sylva et al. study (1999) the relationship between ECERS-R and ECERS-E was ...examined. The correlation coefficient was 0.78 indicating a strong positive relationship between the two measures. Even though the two instruments focus on different dimensions of preschool settings, they both measure a general construct of ‘quality.’ Therefore, it is expected that centers obtaining a high score on the ECERSR will also obtain a high score on the ECERS-E...”

---

<sup>1</sup>The conversion rates used reflect rates as of 11/29/07.

“Apart from the high correlation between the ECERS-E and the ECERS-R, construct validity of this new scale has also been established through the strong relationship with the CIS, a scale for assessing the relationships between setting staff and children.

“Sammons and her colleagues (2002) report significant moderate correlations between the ECERS-E average total and Positive Relationship ( $r = .59$ ) and Detachment ( $r = -.45$ ), two CIS subscales. All correlations were in the expected direction and the correlation coefficients between all the ECERS-E subscales and the CIS subscales ranged from low to moderate, with the positive relationship subscale being moderately associated with all ECERS-E subscales (from .45 to .58)” (Sylva et al., 2003, pp. 44-45).

#### *Predictive validity*

“The predictive validity of the ECERS-E in relation to cognitive progress was found to be better than the power of ECERS-R in the EPPE study on 3,000 children. Controlling for a large number of child, parent, family, home and preschool characteristics, the ECERS-E average total was significantly associated in a positive direction with pre-reading scores, early number concepts and non-verbal reasoning.

“The literacy subscale had a significant positive effect both on pre-reading and on early number concepts. In addition, non-verbal reasoning was significantly affected in a positive direction by the math subscale of the ECERS-E, the diversity subscale and almost significantly by the science and environment subscale. The diversity subscale had also a significant positive effect on early number concepts. As for the behavioral outcomes, although just missing significance at .05, trends of the average total ECERS-E were positive on two of the measures of social/behavioral development: independence/concentration and co-operation/conformity (Sammons et al., in press)” (Sylva et al., 2003, p. 45-46).

#### **Comments**

A number of items in the mathematics and science subscales are optional. For example, when completing the Science subscale, observers would complete items 1 and 2, and then select one of the ‘science activities’ items (3, 4 or 5). This is because, in the fairly limited time observers will spend in a center, observers would not expect to see evidence of the full range of science activities. The choice of optional item is not generally made until later in the observation; observers should gather evidence for all optional items and then score the one for which there is most evidence (i.e. the one which scores the highest).

#### **References**

- Sammons, P., Sylva, K., Melhuish, E., Sijaj-Blatchford, I., Taggart, B., & Elliot, K. (2002). *Measuring the impact of pre-school on children’s cognitive progress over the pre-school period. Technical Paper 8a*. London: Institute of Education.
- Sammons, P., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., & Elliot, K. (2003). *Measuring the impact of pre-school on children’s social behavioural development over the pre-school period. Technical Paper 8b*. London: Institute of Education.

Sylva, K., Siraj-Blatchford, I., Melhuish, E., Sammons, P., Taggart, B., Evans, E., et al. (1999). *Characteristics of the centres in the EPPE sample: Observational profiles. Technical Paper 6*. London: Institute of Education.

---

## Early Language and Literacy Classroom Observation Toolkit (ELLCO)

### Background Information

#### Author/Source

Source: Smith, M.W., Dickinson, D. K., Sangeorge, A., and Anastasopoulos, L. (2002). *Early Language & Literacy Classroom Observation Toolkit: Research Edition*. Baltimore, MD: Paul H. Brookes Publishing Co.

Publisher: Paul H. Brookes Publishing Co.  
Post Office Box 10624  
Baltimore, MD 21285-0624  
Phone: 800-638-3775  
Website: [www.brookespublishing.com](http://www.brookespublishing.com)

#### Purpose of Measure

*As described by the authors:*

“The Early Language and Literacy Classroom Observation (ELLCO) Toolkit... provides researchers and practitioners with a comprehensive set of observation tools for describing the extent to which classrooms provide children optimal support for their language and literacy development. . .

The ELLCO Toolkit is composed of three interdependent research tools. These parts are the *Literacy Environment Checklist*, completed first as a means to become familiar with the organization and contents of the classroom; the *Classroom Observation and Teacher Interview*, used second to gather objective ratings of the quality of the language and literacy environment experiences in a given classroom; and the *Literacy Activities Rating Scale*, completed last to provide summary information on the nature and duration of literacy-related activities observed.” (Smith et al., 2002, p. 1).

#### Population Measure Developed With

“The toolkit has been pilot tested and used in several research studies since its initial development, including research conducted in more than 150 preschool classrooms for the Head Start-funded New England Quality Research Center (NEQRC; 1995-2000) and the Literacy Environment Enrichment Project (LEEP; ongoing), both based in the Center for Children & Families at Education Development Center, Inc., Newton, Massachusetts.

“For the LEEP, the Classroom Observation was used as a pre- and post-intervention measurement tool, with ratings being given in the fall and spring in more than 60 classrooms, including intervention and comparison groups. All of the data come from projects that are concerned with the language and literacy development of children from lower-income families and communities.” (Smith et al., 2002, p. 51).

#### Age Range/Setting Intended For

The ELLCO may be used in PreK to grade 3.

## Ways in which Measure Addresses Diversity

Classroom observation - Item 12, “Recognizing diversity in the classroom,” and Item 13, “Facilitating home support for literacy,” address diversity by measuring the way in which linguistic and cultural diversity are taken into account in classroom activities and conversations, as well as how teachers build on families’ social and cultural experiences. Item 8, “Presence of books,” addresses whether the books in the classroom include representations of various racial and cultural groups. The teacher interview includes a question that gathers information on the teacher’s views of children from diverse racial, ethnic, and language backgrounds.

## Key Constructs and Scoring of Measure

The ELLCO toolkit consists of a literacy environment checklist, a classroom observation component, a teacher interview, and a literacy activities scale.

- *The Literacy Environment Checklist* (24 items) is divided into five conceptual areas:
  - *Book Area* (3 items). Arrangement of classroom’s book area
  - *Book Selection* (4 items). Number, variety, and condition of books in classroom
  - *Book Use* (5 items). Placement and accessibility of books in classroom
  - *Writing Materials* (6 items). Variety of writing tools available for children’s use
  - *Writing Around the Room* (6 items). Evidence of writing activities
- *The Classroom Observation* (14 items) is scored from 1 (*deficient*) to 5 (*exemplary*) and is divided into:
  - *General Classroom Environment*. Organization of the classroom, contents of the classroom, presence and use of technology, opportunities for child choice and initiative, classroom management strategies, classroom climate
  - *Language, Literacy, and Curriculum*. Oral language facilitation, presence of books, approaches to book reading (PreK and kindergarten version), reading instruction (school-age version), approaches to children’s writing (PreK and kindergarten version), writing opportunities and instruction (school-age version), approaches to curriculum integration, recognizing diversity in the classroom, facilitating home support for literacy, approaches to assessment
- *The Teacher Interview*. Consists of questions that help clarify and complete the observation
- *The Literacy Activities Rating Scale*.  
 “[C]onsists of nine questions divided into two categories, Book Reading and Writing. The first three questions gather information on the number of full-group book reading sessions observed, the number of minutes spent in book reading,

and the number of books read. The data for these questions must be recorded in two ways: as amounts...and as scores” (Smith et al., 2002, p. 19).

## Administration of Measure

### Who Administers Measure/Training Required

*Test Administration:* Depending on the purpose of its use, researchers, supervisors, program directors, principals, administrators, and/or teachers may use the ELLCO. It is recommended that potential users have strong background knowledge of children’s language and literacy development, as well as teaching experience in the intended age range.

*Training Required:* A minimum of nine hours of training is required for appropriate and responsible use.

### Setting

The ELLCO may be used in early childhood and early elementary classrooms.

### Time Needed and Cost

*Time:* Approximately 1 – 1 ½ hours.

*Cost:* User’s Guide and Toolkit: \$50.00

## Functioning of Measure

### Reliability Information

#### *Inter-rater Reliability*

- *Literacy Environment Checklist:* When observers have been trained and supervised appropriately, the average inter-rater reliability achieved was 88%.
- *Classroom Observation:* When observers are trained and supervised appropriately, inter-rater reliabilities of 90% and better have been consistently achieved.
- *Literacy Activities Scale:* When observers have been trained and supervised appropriately, the average inter-rater reliability achieved was 81%.

#### *Internal Consistency*

- Literacy Environment Checklist:
  - “Cronbach’s alpha of .84 for the Total score shows good internal consistency. All item-total correlations were moderate to high ( $r = .15$  to  $r = .55$ ). Cronbach’s alpha of .73 for the Books subtotal shows good internal consistency for this composite. All item-total correlations were moderate ( $r = .21$  to  $r = .54$ ) with the exception of Item 1 in the Book Area section (“Is an area set aside just for book reading?”), which exhibited a correlation of .16.

Cronbach's alpha for the Writing subtotal was .75, also indicating somewhat low but still acceptable internal consistency. Item-total correlations ranged from a low of .21 for Item 15 in the Writing Materials section ("Are there templates or tools to help form letters?") to a high of .59 for Item 21 in the Writing Around the Room section ("How many varieties of children's writing are on display in the classroom?") (Smith et al., 2002, pp. 53-54).

- **Classroom Observation:**

"Cronbach's alpha of .83 for the *General Classroom Environment* shows good internal consistency for this composite. All of the item-total correlations were high – with correlation coefficients ranging from .60 for Item 1, Organization of the Classroom, to .75 for Item 6, Classroom Climate – with the exception of Item 2, Contents of the Classroom. This item had the lowest item-total correlation, which was nonetheless a moderate correlation ( $r = .53$ ). The internal consistency of the *Language, Literacy, and Curriculum* composite is very good, with an alpha of .86. All of the item-total correlations were moderate to high, ranging from .55 for Item 8, Presence of Books, to .65 for Item 13, Facilitating Home Support for Literacy. Cronbach's alpha of .90 also shows very good internal consistency for all items combined on the Classroom Observation. All of the item-total correlations for the *Classroom Observation Total* were moderate to high ( $r = .39$  to  $r = .68$ )" (Smith et al., 2002, pp. 57-58).
- **Literacy Activities Rating Scale:**

"Cronbach's alpha of .66 for the *Total* score shows somewhat low but acceptable internal consistency for this measure. Item-total correlations ranged from a low of .17 for Item 9 ("Did an adult model writing?") to a high of .49 for Item 1 ("How many full-group book reading sessions did you observe?"). Cronbach's alpha of .92 for the *Full-Group Book Reading* subtotal shows excellent internal consistency for this composite. All item-total correlations were high ( $r = .79$  to  $r = .88$ ). The Cronbach's alpha for the Writing subtotal was .73, indicating good internal consistency. Item-total correlations were moderate to high, ranging from a low of .37 for Item 9 ("Did an adult model writing?") to a high of .64 for Item 7 ("Did you see children attempting to write letters or words?"). Given the stronger psychometric properties of the two subscales, it is recommended to use the scores on the distinct subscales of the Literacy Activities Rating Scale instead of the total score" (Smith et al., 2002, pp. 62-63).

## Validity Information

### *Criterion Validity*

- **Classroom Observation:**

"The Classroom Observation has been used in correlational research and employed in hierarchical linear modeling designed to determine the contributions of classroom quality to children's receptive vocabulary (Peabody Picture Vocabulary Test-Third Edition; Dunn & Dunn, 1997) and early literacy scores (Profile of Early Literacy Development; Dickinson & Chaney, 1998)...Level-one models examining between-group variability took into account variables such as home language..., gender, and age. The variance in scores that was not accounted for by background factors (15% for



vocabulary, 20% for literacy) was attributed to classroom factors. [The developers'] models examining sources of classroom-related variance found that scores on the Classroom Observation accounted for 80% of the between-classroom variance in vocabulary and 67% of the between-classroom variance in early literacy (Dickinson et al., 2000)" (Smith et al., 2002, pp. 60-61).

#### *Concurrent Validity*

- *Classroom Observation*: Moderate correlations for three Classroom Observation variables with scores on the Assessment Profile for Early Childhood Programs' (Abbott-Shim & Sibley, 1998) Learning Environment subscale:
  - General Classroom Environment subtotal:  $r = .41$
  - Language, Literacy, and Curriculum subtotal:  $r = .31$
  - Classroom Observation Subtotal:  $r = .44$

No relationship was found with the Assessment Profile for Early Childhood Programs' Scheduling subscale (this also "provides divergent validity because the Classroom Observation was developed to tap a construct that is distinct from that examined by the Scheduling subscale") (Smith et al., 2002, p. 60).)

#### *Content Validity*

Experts in the field of early literacy contributed to both the development and the review of the ELLCO Toolkit. Furthermore, all of elements of the ELLCO are aligned with findings presented in *Preventing Reading Difficulties in Young Children* (Snow et al., 1998) and *Learning to Read and Write: Developmentally Appropriate Practices for Young Children* (International Reading Association [IRA] and National Association for the Education of Young Children [NAEYC], 1998).

### **Comments**

A revised version of the ELLCO was made available in early 2008 with a preschool instrument and a separate, more robust measure for use in K-3 classrooms.

### **References**

- Abbott-Shim, M., and Sibley, A. (1998). *Assessment Profile of Early Childhood Programs*. Atlanta, GA: Quality Assist.
- Dickinson, D. K., & Chaney, C. (1998). *Profile for Early Literacy Development*. Newton, MA: Education Development Center, Inc.
- Dickinson, D. K., Sprague, K., Sayer, A., Miller, C., Clark, N., and Wolf, A. (2000). Classroom factors that foster literacy and social development of children from different language

backgrounds. In M. Hopman (Chair), *Dimensions of program quality that foster child development: Reports from five years of the Head Start Quality Research Centers*. Poster presented at the biannual National Head Start Research Conference, Washington, DC.

Dunn, L.M., & Dunn, L.M. (1997). *Peabody Picture Vocabulary Test - 3rd Edition*. Circle Pines, MN: American Guidance Service.

International Reading Association [IRA] and National Association for the Education of Young Children [NAEYC]. (1998, July). Learning to read and write: Developmentally appropriate practices for young children. *Young Children*, 53(4), 30-46.

Snow, C. E., Burns, M. S., and Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, D.C.: National Academy Press

---

## ***Annotated Bibliography***

---

### **ECERS-R**

Aboud, F.E. (2006). Evaluation of an early childhood preschool program in rural Bangladesh. *Early Childhood Research Quarterly*, 21, 46-60.

**Purpose:** The purpose of this study was to evaluate an early childhood program for children between the ages of four-six years of age in rural Bangladesh. Children's cognitive, language, school readiness, and social skills were measured and the quality of the program was evaluated in light of international standards.

**Methods:** Children from preschool programs in three rural sites and a comparison group of children from neighboring villages who did not attend preschool were randomly selected in this cross-sectional study design. This resulted in a total sample size of 401 children, 213 in the preschool group, and 188 in the comparison group, ranging in age from four-and-one-half to six-and-one-half. Instruments used in this study consisted of the three Wechsler Preschool and Primary Scale of Intelligence (WPPSI III) subtests measuring children's cognitive skills, a measure of school readiness, the Play Observation Scale measuring children's social development, parent interviews, and measures of preschool quality, including the ECERS-R, the TECERS (South Indian adaptation of the ECERS), ratings of preschool materials, and a review of preschool program manuals.

**Findings:** Children attending the early childhood programs were found to have higher scores on measures of cognitive development, school readiness, and social play than children in the comparison group. The quality of programs was found to be correlated with child outcomes with higher quality programs resulting in better child outcomes. Since overall quality as measured by the ECERS-R was significantly correlated with group cognitive and school readiness scores, the measure seems to be a valid measure of quality even in an international context.

Bryant, D.M., Clifford, R.M., and Peisner, E.S. (1991). Best practices for beginners: Developmental appropriateness in kindergarten. *American Educational Research Journal*, 28(4), 783-803.

**Purpose:** The purpose of this study was to examine and document the degree to which developmentally appropriate practices are followed in North Carolina's kindergarten classrooms. Furthermore, the authors of this study attempted to examine the factors that

constitute developmentally appropriate kindergarten classes and see whether they matched previous research findings on developmentally appropriate preschool classrooms.

**Methods:** A total of 103 kindergarten classrooms in North Carolina were randomly selected to participate in this study. Each classroom was observed using a modified version of the ECERS (minus the adult needs subscale and modified items reflecting activities specific to kindergarten classrooms) and the Checklist of Kindergarten Activities (CKA), a new measure developed for this study based on NAEYC recommendations for kindergarten practices. All principals and teachers in participating classrooms were surveyed regarding their familiarity with and attitude towards developmentally appropriate practices for kindergarten classrooms.

**Findings:** Sixty percent of classes were found to be below a criterion of developmental appropriateness (a 5.0 or higher mean score on the ECERS), 20% were in a range close to this criterion, and only 20% met or exceeded the criterion. There was wide variation across classrooms within ECERS subscales with a tendency for classrooms with higher scores scoring high on all subscales, and classrooms with lower scores scoring low on most subscales. Cultural awareness, toileting practices, free play, and creative activities were found to be particularly low. The CKA was found to be highly related to the modified ECERS. Teacher and principal scores on developmental appropriateness were the most important predictors of quality and appropriate instruction. Geographic location, per student expenditure, and school size did predict quality. The authors of this study also found that boys and younger children tend to be retained more frequently.

Burchinal, M.R., and Cryer, D. (2003). Diversity, child care quality, and developmental outcomes. *Early Childhood Research Quarterly*, 18, 401-426.

**Purpose:** The purpose of this study was to examine whether child care quality predicts child outcomes for children of color (African-American and Latino) differently than for White children. The authors of this study were also interested in examining the impact of a cultural match between child care provider and children on children's cognitive and social development. Finally, the impact of discrepancies between the caregiver's and the mother's attitudes towards child-rearing on child outcomes was examined.

**Methods:** Data from the Cost, Quality, and Outcomes Project and the NICHD Study of Early Child Care were analyzed to answer the research questions. The child outcome component of the Cost, Quality, and Outcomes Project was based on a subsample of 177 classrooms in 170 centers drawn from a pool of 401 randomly selected full-time

centers in four regions of the United States. Child care quality was measured using the ECERS. Child care provider sensitivity was measured using the Caregiver Interaction Scale (CIS), teaching style was rated using the UCLA Early Childhood Observation Form (ECOF), and child care provider responsiveness was measured using the Adult Involvement Scale (AIS). The measures used in this study were found to be moderately to highly correlated (correlations of .74 to .91 between ECERS, CIS, ECOF; and .26 to .31 between AIS and other measures) and a single composite quality index was calculated. The Peabody Picture Vocabulary Test – Revised (PPVT-R) was used to assess receptive language comprehension and two subtests of the Woodcock-Johnson Tests of Achievement – Revised (WJ-R) were used to measure pre-reading and pre-math skills. Teachers' perceptions of children's social and cognitive skills were measured using the Classroom Behavior Inventory (CBI). Demographic information was collected using a parent questionnaire.

For the second sample of this study, 1,364 families from the total pool of 5,265 families recruited for the NICHD Study of Early Child Care from hospitals after giving birth were enrolled in the study. Child care quality was measured using the Observational Record of the Caregiving Environment (ORCE). The Bracken School Readiness Scale was used to assess school readiness skills, and the Reynell Developmental Language Comprehension Scale (RDLS) was used to assess verbal comprehension. Behavior problems were measured using the Child Behavior Checklist – 2/3 (CBLC), and pro-social behaviors were calculated using the Express and Comply scales of the ASBI.

**Findings:** Standard measures of child care quality were found to be reliable and valid for children across different cultural backgrounds and predicted positive school readiness outcomes for children from different backgrounds. Measures of sensitive and stimulating caregiver behaviors were found to predict school success for White children only, but predicted cognitive outcomes for all children. The NICHD SECC study also found these care-giving behaviors to be related to children's social skills. Reliability, validity, and internal consistency were found to be comparable and good for children from all backgrounds in both studies.

Cassidy, D., Hestenes, L., Hegde, A., Hestenes, S., and Mims, S. (2005). Measurement of quality in preschool child care classrooms: An exploratory and confirmatory factor analysis of the Early Childhood Environmental Rating Scale-Revised. *Early Childhood Research Quarterly*, 20(3), 345-360.

**Purpose:** This study was designed to look at the psychometric properties of the Early Childhood Environment Rating Scale-Revised (ECERS-R), using a large sample.

Specifically, the authors attempted to examine whether there are underlying distinct constructs of quality in the ECERS-R, and whether a shorter version of the scale could reliably be used for research purposes. Unlike previous factor analytic studies, this study uses a large sample size, allowing exploratory and confirmatory analyses to be conducted.

**Methods:** Data were collected from 1,313 preschool classrooms in child care programs across North Carolina from 1999 to 2002. ECERS-R observations were completed only in programs that were striving for a higher star rating in North Carolina's Star Rated License process. Highly trained observers completed the ECERS-R observations during a three-four hour observation period. Teacher interviews (lasting about 30 minutes) were conducted at the end of the observations to clarify information and complete unobservable items. Information on teachers (such as level of education), group size, and teacher/child ratios was also collected. Descriptive statistics were computed for each item, subscale, and overall scores to assess for normality and missing data.

**Findings:** A shorter version of the ECERS-R was established, which uses 16 items from the original instrument and can be broken into two subscales: Activities/Materials and Language/Interaction. These subscales performed well as a proxy for the ECERS-R. The new 16-item assessment is highly correlated with total ECERS-R scores ( $r = .90$ ) and successfully differentiated high-quality from low-quality programs. It should be noted that all classrooms in this study were sampled from high-quality programs; these findings may not hold true in lower-quality programs.

Dowda, M., Pate, R.R., Trost, S.G., Almeida, M.J., and Sirard, J.R. (2004). Influences of preschool policies and practices on children's physical activity. *Journal of Community Health*, 29(3), 183-196.

**Purpose:** Given recent concerns about obesity and overweight in preschool and school-age children, the purpose of this study was to examine whether and how the amount of physical activity to which preschool children are exposed is influenced by preschool program quality and policies/practices regulating the program.

**Methods:** Nine private, church-related, and Head Start preschool programs (three of each type) were randomly selected from a pool of preschools in Columbia, South Carolina. Within these programs, 277 children participated in this study. Children's height and weight were measured and a body mass index (BMI) was calculated for each child. Levels of physical activity were observed using the Observation System for Recording Activity in Preschools (OSRAP), a modified version of the Children's Activity

Rating Scale (CARS). Overall quality was observed in one classroom per preschool using the ECERS-R. Information on school policies and practices related to physical activity was collected via a structured administrator interview. A parent questionnaire was used to gather basic demographic information on children participating in the study.

**Findings:** No differences were found between the three different types of preschool programs in the amount of moderate to vigorous physical activity in which children were engaged. Different school policies and practices related to physical activity, however, were significantly related to physical activity levels. Programs that employed teachers who had completed college and programs that offered regular field trips were shown to have particularly high levels of moderate to vigorous physical activity and provided more time on the playground. Children in programs that had higher quality scores on the ECERS-R spent less time in sedentary activities. Overall, children were found to spend less than the recommended time in physical activity.

Goelman, H., Forer, B., Kershaw, P., Doherty, G., Lero, D., and LaGrange, A. (2006). Towards a predictive model of quality in Canadian child care centers. *Early Childhood Research Quarterly*, 21, 280-295.

**Purpose:** This paper uses data from an extensive study of quality in Canadian child care centers to examine predictors of quality. The goals of the study were to identify the proximal and distal factors that affect child care quality, to examine the direct and indirect relations among these factors and quality, and to do so in a national sample of child care centers in Canada so as to complement, and perhaps also contrast with, the large national studies of child care in the United States. This study permits examination of whether patterns identified in child care in the United States, such as ratio, professional development, and turnover, hold for Canadian child care.

**Methods:** Data for this study are from the second phase of the Canadian You Bet I Care Project (YBIC). Data on this phase of the project are drawn from seven jurisdictions chosen to reflect a range of child care policies and regulations: Alberta, British Columbia, New Brunswick, Ontario, Quebec, Saskatchewan, and Yukon Territory. Two hundred thirty nine centers participated; 108 with infant/toddler as well as preschool classrooms, 117 with preschool classes only, and 14 with infant/toddler classrooms only. Observations of quality were carried out in one classroom per age group within each setting so that only one classroom was observed in settings with only one age group. A questionnaire was used to collect information on children served, staff background and wages, center policies, and center revenues and expenditures. Three measures of

quality were collected: the Caregiver Interaction Scale (CIS); the Infant/Toddler Environment Rating Scale (ITERS); and the Early Childhood Environment Rating Scale-Revised (ECERS-R).

**Findings:** For both the ITERS and ECERS-R, subscales were all significantly intercorrelated (often highly intercorrelated) with each other and with the total score. As a result, analyses focused only on total scores. In all jurisdictions, ITERS and ECER-R total scores were higher for non-profit than for commercial child care settings. Quality total scores were slightly higher for preschool classrooms than for infant/toddler classrooms, with 44 percent of preschool classrooms being rated as “good” or higher, compared to that rating for only 29 percent of infant/toddler classrooms. The distribution of ITERS and ECERS-R scores was approximately normal. Total scores on the ITERS and ECERS-R varied significantly by jurisdiction. Path analyses were used to examine direct and indirect pathways to observed quality. For both infant/toddler classrooms and preschool classrooms, a set of direct predictors of quality is strengthened by a set of indirect predictors. For infant/toddler classrooms:

- Direct predictors of quality included number of adults in the classroom and teacher education level.
- Indirect predictors of quality (related to quality through the direct predictors) included parent fees, adult-to-child ratio, and whether the center was used as a practicum site for student teachers.

For preschool classrooms:

- Direct predictors of quality included staff wages, teacher education, number of staff in the classroom, staff satisfaction with co-workers and the work environment, and subsidized or free rent and/or utilities.
- Indirect predictors of quality (related to quality through the direct predictors) included: operating auspice; adult-to- child ratio; parent fees; and number of staff in the room.

The authors note that “despite differences in child care regulations, the overall levels of child care quality were disturbingly low in child care centers in both the US and Canada with most ITERS and ECERS scores in the ‘minimal’ range...and rarely reaching the ‘good’ to ‘excellent’ ...threshold.” (p. 293).



Perlman, M., Zellman, G.L., & Le, V. (2004). Examining the psychometric properties of the Early Childhood Environment Rating Scale-Revised (ECERS-R). *Early Childhood Research Quarterly*, 19(3), 398-412

**Purpose:** The purpose of this study was to examine the ECERS-R by assessing its key psychometric properties. By doing so, the authors sought to gain insight into the reliability of the ECERS-R. Administration of shorter versions of the ECERS-R were evaluated and found to be advantageous.

**Methods:** Staff of the Center for Human Investment Policy (CHIP) at the University of Colorado, Denver, collected ECERS-R data from 326 classrooms in 202 child care centers in Colorado. This data collection took place from Fall 2000 through Spring 2002. The ECERS-R data collectors were experienced early childhood education practitioners and held an associate's degree or higher. They were required to achieve reliability on three consecutive ECERS-R administrations with a percent agreement of 85% or higher on each item. Caregiver-to-child ratios were obtained at selected times each day for four consecutive weeks. Staff credentials for the lead teacher in each classroom were gathered through staff surveys and included level of education, teaching experience, and early childhood education training.

**Findings:** The average ECERS-R score for the sample was 5.153, falling within the "very good" range. The authors hypothesize that this high rating could reflect the fact that all classrooms had previously been involved in quality improvement initiatives and many had already been assessed using the ECERS and ECERS-R. The correlations among individual items as well as between subscales were found to be fairly high (medians of .63 and .62 respectively). The internal consistency estimate, which indicates the degree to which the items measure the same construct, was .95. All of these statistics suggest that the ECERS-R could be conceptualized as measuring fewer than seven distinct aspects of quality.

A factor analysis was conducted to determine how many distinct aspects of quality were actually being measured by the ECERS-R. This analysis confirmed that only three factors were retained. Although three were found, it is important to note that all three aspects of quality were highly related, further suggesting that the ECERS-R is a global assessment of child-care quality. To further test this, the authors randomly chose three subsets of 12 items from the ECERS-R and found that the total ECERS-R score as well as each of the three randomly selected subsets of items were highly correlated with caregivers' years of experience and not significantly correlated with teachers' educational attainment, post-secondary early childhood education coursework, or child-

to-staff ratio. One implication of these results is that shorter versions of the ECERS-R could reliably measure global quality.

Sakai, L.M., Whitebook, M., Wishard, A., and 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.

**Purpose:** “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...” (p. 428). The most widely used measure of overall quality prior to 1998 was the Early Childhood Environment Rating Scale (ECERS). In 1998, a revision of the ECERS was made available. The revision sought to respond to changes in understanding of best practices for young children, as well as to address issues that had arisen in the use of the ECERS. The authors conducted three focus groups with researchers and practitioners addressing the cultural sensitivity of the ECERS as well as its functioning for inclusion of children with disabilities. They also carried out content analysis of the ECERS and other measures of quality and reviewed research conducted using the ECERS. The revised ECERS: no longer includes infant/toddler items, which are now contained in the Infant/Toddler Environment Rating Scale; eliminates items considered to be redundant; includes new areas, such as use of computers and videos; add indicators and examples to improve its focus on inclusiveness and cultural sensitivity; and makes changes in formatting and scoring to increase consistency with the ITERS and Family Day Care Rating Scales. Field tests of the revised measure by its authors found high inter-rater reliability and reasonable internal consistency at the level of total score and subscales. The authors of the present study used both the original and revised measure in the third round of a longitudinal study of change over time in quality as child care centers worked towards or maintained NAEYC accreditation to examine the equivalence of the original and revised measures.

**Methods:** The sample included 68 classrooms in 43 child care centers participating in the longitudinal study of NAEYC accreditation in northern California. Both the ECERS and ECERS-R were completed in each classroom along with the Caregiver Interaction Scale. Lead teachers in each classroom and center directors were interviewed.

**Findings:** The distribution of ratings on the total score for the original and revised ECERS were quite similar, and within-classroom scores on the two measures were highly correlated. Factor analysis with the original measure had indicated two factors, with the one focusing more on materials, schedules, and activities, and the other

focusing more on adult-child interactions, supervision, and discipline. Principal component factor analysis on the revised measure indicated that a two-factor solution also had the best fit. The two factors, though, are somewhat less clearly differentiated according to tone of interactions and provision of materials for learning, with some items that had loaded on each factor for the original measure not doing so in the reviewed version. Internal consistency reliability was high, both for the total score and two subscales based on the factor analysis. With few exceptions, the items on each subscale correlated more highly with a summary score for that subscale than the other subscale. For both the original and revised measures, classrooms with total scores of good or higher overall differed significantly from those with lower overall scores on average Caregiver Interaction Scale sensitivity and harshness subscale scores. Classrooms overall scored low on the items related to cultural awareness on both the original and revised measure, and no differences were found on the global score or cultural awareness item according to language match of teachers and children when children spoke a language other than English. Global quality scores on both the original and revised measure did not differ significantly according to whether teachers indicated concerns about supports and resources for children with special needs. Further, there was no relationship between global score on either the original or revised measure and retention of staff with a Bachelor's degree or advanced training in early childhood education. The results indicate that the original and revised ECERS are comparable as measures of quality. The authors note that as the ECERS and ECERS-R come to be used increasingly as tools for self assessment, it is possible that "grade inflation" may be taking place, with programs purchasing the materials and carrying out specific procedures, like hygiene practices, that are related to higher scores. While these intentional changes are all positive, such "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 that demand more depth of understanding of child development and exposure to skilled role models." (p. 443).

Warash, B. G., Markstrom, C.A., and Lucci, B. (2005). The Early Childhood Environment Rating Scale-Revised as a tool to improve child care centers. *Education*, 126 (2), 240-250.

**Purpose:** The evidence indicates that many children in the United States participate in child care that is not of high quality. States are utilizing different strategies to strengthen child care quality, including providing incentives for going beyond basic levels of quality, enforcing tougher regulatory standards, clarifying the steps for improving professional development through career lattices, providing training for early childhood educators as

well as directors of early childhood programs on developmentally appropriate practice (DAP), and providing technical assistance to directors of early childhood programs for improving the quality of their programs. The present study examined the use of the Early Childhood Environment Rating Scale-Revised (ECERS-R) as a tool in technical assistance aimed at improving quality. In particular, the study examined whether quality would improve (overall and on specific subscales) when initial scores on the ECERS-R, as well as a quality improvement plan, were reviewed with child care center directors, who then provided this information to teachers.

**Methods:** Educare was an initiative established in West Virginia in 2000 to improve the preschool experiences of young children. One measure used in the state evaluation of Educare was the ECERS-R. The present study used the ECERS-R scores of four participating child care centers, with eight classrooms serving children between the ages of three and five, in one county. Classroom observations using the ECERS-R were completed by pairs of observers, and any discrepancies in ratings were discussed and resolved while they were still in the classroom. After completing the initial ECERS-R, a training plan was developed for each classroom using the information from the ECERS-R ratings. The training plan, along with a summary of recommended improvements, was shared with the center director during a two- to three-hour meeting. Directors, in turn, shared the information with the classroom teachers. The training plan noted scores on each of the 43 indicators in the ECERS-R, along with guidance on how to bring the scores up one level. Some of the recommended steps for improvement were specific and concrete (for example, provide opportunities for indoor play with sand), while others were broad and required further training (such as preventing conflict among children). A second observation using the ECERS-R was conducted 7-10 months after the initial observation. Analyses involved comparing pre- and post-test scores on the seven ECERS-R subscales, as well as on the total score.

**Findings:** A significant improvement from pre- to post-test occurred for the ECERS-R total score as well as on three of the seven subscale scores: personal care routines, activities, and interactions. Improvements on two further subscales approached statistical significance: program structure, and parents and staff. The authors conclude that the ECERS-R can be an effective tool for working with center directors to guide quality improvement. They underscore the importance of articulating specific quality improvement objectives for each of the ECERS-R indicators.

Warash, B.G., Ward, C. and Rotilie, S. (2008). An exploratory study of the application of Early Childhood Environment Rating Scale criteria. *Education*, 128 (4), 645-658.

**Purpose:** West Virginia passed legislation in 2002 calling for full implementation of a voluntary pre-kindergarten program for all four-year-olds by 2012/13. The legislation calls for school districts to contract for at least 50% of their pre-kindergarten services in existing community-based programs, such as child care and Head Start. Annual plans for increased access are developed at the county level by collaborative groups that must include representation from public schools (preschool program and special needs program), Head Start, licensed child care, and private nursery schools. The planning process includes a focus on quality according to policy requirements specified in Policy 2525 for such features as group size and ratio, use of curricula, professional development for staff, parent involvement, and health and safety. This policy also requires an annual classroom observation for each pre-kindergarten classroom using the ECERS-R. Each county submits an annual plan reporting on ECERS-R scores and plans for improvement. A three-module training sequence has been developed by West Virginia early childhood professionals aimed at developing a group of reliable ECERS-R observers. The first module, which takes six hours and focuses on developing an understanding of the ECERS-R structure, purpose, items, scoring, and links with state-approved curricula, is available for staff development, and can be used towards an annual 15-hour training requirement focusing on curriculum and inclusion. The purpose of the present study was to describe changes that teachers reported making in their classrooms after completing this module of training on the ECERS-R.

**Methods:** Six months after participating in the first module of the ECERS-R training provided for West Virginia teachers working in pre-kindergarten classrooms, teachers were asked to complete an e-mailed questionnaire. The training was given by one of the developers of the ECERS-R, Thelma Harms. The questionnaire was e-mailed to 35 participants in the Module One training, and 11 teachers responded. Respondents indicated changes that they had made with respect to each of the subscales and scoring items from the ECERS-R. They were also asked to provide information on the characteristics of their classrooms.

**Findings:** Respondents provided pre-kindergarten to children between the ages of four and five in a range of settings, including Head Start, child care, public school, or a combination of these. Items on which teachers most often indicated making changes fell in the following subscales:

- Space and Furnishing,
- Personal Care Routine,
- Activities, and
- Program Structure

Most of the items for which change was reported involved Space and Furnishings. The authors note that changes to the physical environment may be easier to accomplish than other changes. Specific items from other subscales that were changed by a substantial proportion of respondents included the schedule item of the Program and Structure subscale (with respondents reporting increasing the amount of time for free play); the item regarding promoting acceptance for diversity in the Activities subscale (with respondents reporting that they added more multicultural dolls, books and puzzles to the classroom); and the health practices item in the Personal Care and Routine subscale (with respondents reporting establishing hand-washing routines and increasing the emphasis on hygiene). Subscales in which changes were reported less frequently were Language and Reasoning, Interactions, and Parents and Staff. The authors note that this study involved a small sample and self-reporting rather than observational data. Nevertheless, the findings suggest that even a relatively brief training focusing on the ECERS-R as a tool to describe the specific facets of quality in early childhood settings has the potential to help teachers begin to make small positive changes in their classrooms.

---

## CLASS

Burchinal, M., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., and Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher-child interactions and instruction. *Applied Developmental Science, 12* (3). 140-153.

**Purpose:** Nearly three-fourths of states now have pre-kindergarten programs, and these programs serve more than 25 percent of four-year-olds in this country. An underlying assumption is that high-quality early childhood programs will support positive child outcomes, yet there has been limited examination of the assumption that pre-kindergarten programs are of high quality or that they support children's development in the transition to formal schooling. The aims of this study were to examine (1) specific aspects of classroom quality in pre-kindergarten programs in six states; (2) the development of children in these programs both during the pre-kindergarten year and the kindergarten year; and (3) the associations between specific aspects of classroom

quality and children's development. Earlier research reported by these investigators found that the observed quality of pre-kindergarten programs was related to gains in children's development during the pre-kindergarten year. The present study asks whether the link between the quality of pre-kindergarten and child outcomes is maintained into the kindergarten year.

**Methods:** A stratified random sample of 40 state funded pre-kindergarten sites was selected in each of six states with mature pre-kindergarten programs. The sample within each state was stratified with respect to location of the pre-kindergarten in a school or community setting, full vs. part-day programs, and whether teachers were required to have a Bachelor's degree. One classroom within each site was randomly selected. Four children in these classrooms were randomly selected from among those whose parents provided consent, who would be old enough to attend kindergarten the next year, did not have an Individualized Education Program (IEP), and spoke English or Spanish well enough to follow simple instructions according to the teacher. There were seven hundred forty six children for whom data were available on classroom quality in both the pre-kindergarten and kindergarten year as well as on the children's development (although data on the children's development for fall and spring of both the pre-kindergarten and kindergarten year were not always complete). The children were from families with varied parental education, resources, race, and ethnicity, although a majority were from low-income families and 43 percent were White/non-Hispanic. Two measures of classroom quality were collected: The Early Childhood Environment Rating Scale-Revised (ECERS-R) and the Classroom Assessment Scoring System (CLASS). Both measures were collected in both the fall and spring during the pre-kindergarten year. During the kindergarten year, the ECERS-R was collected in the winter, and the CLASS was collected at least once in the winter or spring. Factor analyses conducted with the ECERS-R data indicated two factors. The present analyses examined the first factor, labeled *Interactions and Teaching*, which included indicators of staff-child interactions, discipline practices, supervision, encouraging the children to communicate, and using language to develop reasoning skills. Factor analyses with the CLASS data also indicated two factors. The *Instructional Climate* factor, which is the focus in the present analyses, included the CLASS dimensions of Concept Development and Quality of Feedback. Children's outcomes included direct assessments using standardized measures of academic achievement and teacher ratings of the children's social skills. Composite scores were created to reflect the children's language, reading, and behavioral adjustment.

**Findings:** Mean scores on the ECERS-R Interactions and Teaching factor were in the medium range, while those for the CLASS Instructional Climate factor were in the low range. Both positive interactions with the pre-kindergarten teacher (the ECERS-R factor) and the quality of instruction in the pre-kindergarten classroom (the CLASS factor) predicted language and reading outcomes in kindergarten, controlling for scores at the start of pre-kindergarten as well as background characteristics. Children whose mothers had less education were rated as having stronger social skills and fewer behavior problems if they had attended pre-kindergarten programs for more hours in which observed quality was higher on the ECERS-R factor involving positive interactions between teachers and children. In sum, “the instructional quality of the pre-k program significantly, though modestly, predicted language, academic, and social performance of children up to one year beyond PK” and “[i]t is noteworthy that both positive interactions with the PK teacher and the instructional quality of the PK classroom were important.” While pointing to enduring associations of pre-kindergarten quality and child outcomes, the findings raise concerns about the average quality of pre-kindergarten programs.

Early, D. M., Bryant, D. M., Pianta, R. C., Clifford, R. M., Burchinal, M. R., Ritchie, S., Howes, C., and Barbarin, O. (2006). Are teachers' education, major, and credentials related to classroom quality and children's academic gains in pre-kindergarten? *Early Childhood Research Quarterly, 21*. 174-195.

**Purpose:** This study describes associations between levels and forms of teachers' education, major, and credentials and their relationship to classroom quality and children's academic gains. Although there is widespread agreement that teachers' education and training are important, the field lacks common ways of approaching the measurement of these constructs.

**Methods:** Six states were selected from those that had committed significant resources to pre-school initiatives in 2001. Random samples were taken from state-funded pre-kindergarten schools and centers (54% in public schools and 49% in part-day programs). A total of 237 lead teachers and 939 children participated in the study. In each classroom, the lead teacher was asked to complete a questionnaire involving education level, major, and credentialing in the fall and spring. Classroom quality was measured with the ECERS-R and the CLASS tools, also in the fall and spring; children were assessed by the PPVT-III, OWLS oral expression scale, Woodcock-Johnson III tests of achievement, Identifying Letters, Identifying Numbers, and Identifying Colors.

**Findings:** Teachers with education beyond the Bachelor's degree scored higher on the ECERS-R Teaching and Interaction subscale than teachers with an Associate's degree.



No other associations were found between teacher education and quality measures. These results contradict much of the previous early childhood research linking teachers' education to higher quality and children's educational gains. Teachers' education level and children's math skills (measured by Woodcock-Johnson Applied Problems) were found to be significantly positively correlated. This was the only association found between teachers' education, major, or credentials and any other early academic skill.

Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., and Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-Kindergarten programs. *Early Childhood Research Quarterly, 23*, 27-50.

**Purpose:** This paper examined children's growth in school-related learning and social skills during the preK year in state-funded programs designed to prepare children for kindergarten. The authors expected this growth to be attributed to dimensions of program quality.

**Methods:** Data for this analysis came from the National Center for Early Development and Learning (NCEDL) Multi-State Study of Pre-Kindergarten and the State-Wide Early Education Programs Study (SWEEP). Classroom observations and child assessments were conducted in 692 randomly selected preK classrooms. Child outcomes were measured both in the fall and spring of the pre-Kindergarten year, along with observations of the classroom environment using the ECERS-R and the CLASS tools.

Researchers used a battery of various child assessments and teacher ratings to assess outcomes in language and literacy, math, and social skills. Direct child assessments of language and literacy were performed using the Peabody Picture Vocabulary Test (PPVT), the Oral and Written Language Scale (OWLS), and the Identifying Letters measure. Teacher ratings of children's language and literacy skills were assessed using items from the Early Childhood Longitudinal Study–Kindergarten Cohort teacher questionnaire. Children's math skills were assessed directly using the *Woodcock-Johnson III Tests of Achievement: Applied Problems Subtest*. Teachers' ratings of children's social skills were assessed using the Social Skills and Behavior Problems scale.

**Findings:** Overall, associations were modest. Structural measures of quality were neither highly correlated with each other nor with measures of process quality, teacher-child closeness, or classroom practices. CLASS Instructional Climate was the only significant predictor of gains in either receptive or expressive language. As a cautionary note, the interval between assessments (six months) may have been insufficient to find

large gains. Also, further work that includes classrooms at the higher end of instructional quality is needed.

La Paro, K.M, Pianta, R.C., and Stuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the prekindergarten year. *The Elementary School Journal*, 104(5), 409-426.

**Purpose:** This article describes the development of the Classroom Assessment Scoring System, in addition to providing validity and reliability data collected after utilizing the measure in a multistate study targeting more than 200 preschool classrooms. The CLASS was developed to contrast with other measures that focus primarily on the physical space and materials in the classroom, rather than teacher-child relationships. We note that there have been revisions to the CLASS since the publication of this article, so that the dimensions measured in the most recent version of the CLASS do not correspond completely to those described here.

**Methods:** Participants included 240 Pre-K programs in six states (40 programs in each state) chosen to reflect a diversity of teacher training, location, length of program, and state funding. One classroom from each center/school was selected to be the target for observation. In each classroom, researchers randomly selected four children to participate. All participating children spoke English or Spanish well enough to understand teacher instructions, did not have an Individualized Education Program (IEP), and had parental consent to participate in the study. Ideally, researchers selected two girls and two boys in each classroom. Four data collectors in each state evaluated the classrooms using the CLASS, the ECERS, and the Snapshot.

**Findings:** The study found that correlations among the nine CLASS scales ranged from .84 to -.03. Positive climate and sensitivity had the strongest positive correlation. Concept development and over-control had the strongest negative correlation. A factor analysis revealed that the majority of the variance was accounted for by two factors: emotional support and instructional support. High positive climate, teacher sensitivity, and behavior management and low negative climate and over-control characterized positive emotional support. High ratings on productivity, concept development, learning formats, and quality of feedback characterized positive instructional support. The ratings from the CLASS were then compared to the ratings from the other measures used to observe classrooms. The CLASS factor scores of emotional support and instructional support were moderately related to the ECERS total score. Emotional support and instructional support factors were most strongly related to the language reasoning and interaction subscales from the ECERS. ECERS measures of program structure,

furnishings, space, and activities were not as strongly related to any individual constructs from the CLASS. Classrooms with high emotional support were coded as having more scaffolding and low levels of didactic teacher-child engagement using the Snapshot. Further, as would be expected, negative climate and over-control, two CLASS scales, were moderately inversely related to children's engagement on the Snapshot.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., Burchinal, M., Early, D. M., and Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*(3), 732-749.

**Purpose:** This study sought to examine how measures of pre-K program quality predicted four-year-old children's academic, language, and social skills through ECERS-R and CLASS observations. Results obtained could further the development and improvement of existing programs and encourage new research.

**Methods:** Participants included 2,439 four-year-old children from 671 preschool classrooms in 11 states. The preschool programs included in the study all had been in existence for several years, were stable, and served a majority of four-year-old children in the state. Classroom quality was measured by three facets: adherence to standards of quality set forth by the National Institute for Early Education Research (NIEER), observations of overall classroom quality using the Early Childhood Environment Rating Scale-Revised (ECERS-R), and observations of teachers' interactions with children and the emotional support provided in the classroom using the Classroom Assessment Scoring System (CLASS). Multiple child assessments were administered, including the Peabody Picture Vocabulary Test (PPVT), the Oral Expression Scale from the Oral and Written Language Scale (OWLS), and subscales from the Woodcock-Johnson-III Test of Achievement. Teachers rated children's social competence and problem behaviors using the Teacher-Child Rating Scale (TCRS) in both the fall and spring.

**Findings:** Controlling for prior skill levels, family and child characteristics, and program characteristics, quality of instructional interactions between teachers and students (measured using the CLASS) was significantly positively correlated with all measures of academic and language development. The quality of emotional support provided by teachers, also measured using the CLASS, was significantly positively correlated with children's development of social skills and significantly negatively correlated with children's problem behaviors, such as class disruption, anxiety, and difficulty following directions. The ECERS-R was positively associated only with children's development of

expressive language. The standards set by NIEER were not consistently associated with measures of academic, language, or social development.

Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., and Barbarin, O. (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.

**Purpose:** The purpose of this study was to examine the extent to which program, classroom, and teacher attributes predict observed quality and teacher-child interactions.

**Methods:** This study took place in six states that were selected from those which had committed significant resources to preK initiatives in 2001. Stratified random samples of centers and schools were taken, after which one classroom in each center/school was selected. Quality was observed using the CLASS, ECERS-R, and Snapshot, and teachers completed questionnaires measuring their attitudes, beliefs, and depressive feelings. Multivariate analyses were then conducted to determine the extent to which the characteristics of the program and teachers predicted quality.

**Findings:** Classrooms in which at least 60 percent of the children were from low-income families were rated significantly lower on the quality measures of Teaching and Interactions and Provisions for Learning from the ECERS-R. Also, teacher characteristics significantly predicted the CLASS Emotional Climate, ECERS-R Interactions, and the ECERS-R Provisions scores. Overall, the findings indicate that program and teacher attributes are statistically significant, although quite modest, predictors of observed quality in preK classrooms. As a cautionary note, the authors assert that the detected associations were very modest.

---

## ECERS-E

Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., and Totsika, V. (2006). Capturing Quality in Early Childhood through Environmental Rating Scales. *Early Childhood Research Quarterly*, 21(1), 76-92.

**Purpose:** The purpose of this study was to introduce the ECERS-E by exploring its relationship to already established measures of quality and investigate its ability to predict child outcomes during the pre-school period in England.

**Methods:** More than 3,000 children from 141 British pre-schools were followed from age three to age seven as part of the Effective Provision of Pre-school Education (EPPE), a large-scale prospective longitudinal study. Five regions in England were strategically

sampled to include socially and ethnically diverse areas. Twenty-five pre-school centers were then randomly selected from each region. The sample of children was close to equally split between girls and boys. Though primarily of White, British heritage, the sample did include children of Black Caribbean, Pakistani, Black African, and mixed ethnic heritages. At the time of entry into the study, the children ranged in age from 35 to 52 months with a mean of three years, four months.

The ECERS-R and the ECERS-E were administered upon entry into the EPPE study. Also, assessments of children's cognitive abilities and social/behavioral competencies were conducted at age three and age five, upon entry to primary school. This paper reported the findings related to child developmental outcomes at age five; follow-up results from ages seven and eleven are planned.

**Findings:** Based on the data from the 141 pre-schools, a significant strong relationship was found between the ECERS-R and ECERS-E total scores, supporting the construct validity of the ECERS-E. After controlling for age and child/family background, the ECERS-E was a significant predictor of children's scores on Pre-reading, General Mathematics Concepts, and Non-verbal Reasoning. The ECERS-E was not predictive of Spatial Awareness or Language measures. With one exception – the interaction scale being predictive of General Math Concepts–the ECERS-R was not a significant predictor of child pre-academic outcomes. The opposite results were found for social/behavioral development; the ECERS-R total score significantly predicted scores on Cooperation and Conformity and the ECERS-E failed to predict any social-emotional outcomes. These findings indicate that the ECERS-R is more sensitive to aspects of quality related to children's social development, whereas the ECERS-E is more sensitive to emerging academic skills.

Sylva, K., Taggart, B., Siraj-Blatchford, I., Totsika, V., Ereky-Stevens, K., Gilden, R., and Bell, D. (2007). Curricular quality and day-to-day learning activities in pre-school. *International Journal of Early Years Education*, 15(1), 49-65.

**Purpose:** The purpose of this paper was to determine how curricular quality, as measured by the ECERS-E, was related to the day-to-day activities and pedagogical activities of teachers experienced by children in pre-school.

**Methods:** Data for this study were drawn from the Effective Provision of Pre-school Education (EPPE) project and its sub-study, the Researching Effective Pedagogy in the Early Years (REPEY) project. From the 141 centers in the EPPE project, ten effective centers with positive child outcomes were chosen for study. These ten centers were

then separated (based on their ECERS-E total quality score) into centers of “adequate” quality and “good” quality. Researchers then randomly selected children in each center to observe throughout the day, coding for curriculum, social grouping, learning activity, and staff-child interactions.

**Findings:** Children in the “good” quality centers spent a significantly greater proportion of time in sustained shared thinking with staff and experiencing direct teaching. Children in “adequate” quality centers experienced significantly more monitoring time, or time in which staff simply monitored the children without interacting. Another difference was that children in “good” quality centers spent more time experiencing Communication, Language, and Literacy activities, and a specific curriculum, while children in the “adequate” quality centers spent more time experiencing Physical Development and Creative Development. Also, in “good” quality pre-schools, children spent more time participating in reading/writing/listening and adult-led activities than in centers of “adequate” quality. There was no difference in amount of time spent in Personal, Social, and Emotional development across the two levels of quality.

---

## ELLCO

Diamond, K. E., Gerde, H. K., and Powell, D. R. (2008). Development in early literacy skills during the pre-kindergarten year in Head Start: Relations between growth in children's writing and understanding of letters. *Early Childhood Research Quarterly*, 23, 467-478.

**Purpose:** This article discusses the connection between children’s writing skills and their knowledge of letter names and sounds. It also examines the relationship between classroom literacy environment and the development of early writing skills. The researchers assert that the importance of writing skills is downplayed in preK research, although this is an integral part of children’s early literacy skills.

**Methods:** The study sample was taken from 35 Head Start classrooms previously participating in a research study examining literacy practices. The mean age for children participating in the study was 53 months, or approximately four-and-one-half years. Approximately half of the study participants were considered to live in urban areas while the other half lived in areas considered to be rural or suburban. All participating children were directly assessed in four areas: writing, alphabet knowledge, initial sounds, and print concepts. In addition, the classroom literacy environment was assessed using only the ELLCO Literacy Environment Checklist. Researchers limited their analyses to the 13 ELLCO Literacy Environment Checklist items that directly assess the availability and use

of writing materials in the classroom. High internal consistency of the ELLCO was documented in this sample ( $\alpha=.93$ ).

**Findings:** The researchers found significant positive associations between children's writing skills and their knowledge of the alphabet. However classroom supports for children's writing, as they were measured with the writing components of the ELLCO Literacy Environment Checklist, were not found to be significantly associated with children's growth in writing over the pre-kindergarten year. It is noted that the ELLCO Checklist records the presence of writing materials rather than documenting writing support and instruction. The authors conclude that although environmental supports for writing are important, it is unlikely that writing materials alone, without supports for their use, can develop children's writing skills.

Dickinson, D. K., and Caswell, L. (2007). Building support for language and early literacy in preschool classrooms through in-service professional development: Effects of the Literacy Environment Enrichment Program (LEEP). *Early Childhood Research Quarterly*, 22, 243-260.

**Purpose:** The purpose of the study was to examine the effects of teacher participation in the Literacy Environment Enrichment Program (LEEP). The LEEP program was designed specifically to improve the quality of language and literacy practices in Head Start classrooms. Previous research has shown that classrooms in low-income areas often have lower scores on the Language, Literacy, and Curriculum subscale of the ELLCO, but have higher scores on the General Classroom Environment. Researchers hypothesized that teachers participating in the LEEP program would should significant improvement in language and literacy scores as compared to the control group teachers.

**Methods:** Teachers were recruited for participation in the LEEP program through calls to Head Start program directors throughout New England asking for recommendations for one or more teaching teams to participate in the program. In total, 70 teachers participated in the study: 30 in the LEEP program group and 40 in the comparison group. Most teachers participating in the study had either an Associate's or a Bachelor's degree. Classroom observations were performed twice during the school year: in the fall prior to the beginning of the LEEP course and in the spring after the end of the LEEP course. The classroom observations consisted of the complete ELLCO toolkit as well as two subscales of the Assessment Profile. High inter-rater reliability on both observational measures was documented. Additionally, prior findings showed internal consistency to be sound on the Classroom Observation and Teacher Interview parts of the toolkit.

**Findings:** The researchers found overall gains in observational scores from fall to spring for both treatment and comparison groups. In addition, *t*-tests on spring scores showed significant differences between treatment teachers and comparison teachers overall and on subscale scores except for the Writing subscale of the Literacy Activities Rating Scale. A moderate to large effect of the intervention was found on the overall scores for the Classroom Observation Scale, the Literacy Environment Checklist, the Literacy Activities Rating Scale, and the Learning Environment subscale of the Assessment Profile. Large to moderate effects were also found on subscale scores for each of the three ELLCO components, as well as for the Assessment Profile. The researchers note the strongest effects of the intervention on scales aimed specifically at measuring the language and literacy environment, a finding consistent with expectations because the intervention was aimed at improving literacy practices rather than general classroom practices.

While there were variable changes in literacy-related practices, there were large gains in Literacy Environment items. A possible explanation for this is that it is easy for teachers to introduce literacy-related materials into their classrooms, but it can be much more difficult for teachers to change their practices to provide support for engaging with the new materials in the room. The researchers also point out that writing in the early childhood classroom could be a new aspect of language and literacy instruction for some teachers. They conclude that obtaining a better understanding of how professional development affects classroom practices requires more fine-grained measures of classroom activities and practices.

Gettinger, M., and Stoiber, K. (2007). Applying a response-to-intervention model for early literacy development in low-income children. *Topics in Early Childhood Special Education, 27*(4), 198-213.

**Purpose:** This article details the implementation and evaluation of the Exemplary Model of Early Reading Growth and Excellence (EMERGE) program in fifteen classrooms throughout Milwaukee, Wisconsin. The EMERGE program is an Early Reading First project funded through the U.S. Department of Education. The program is aimed at helping low-income children obtain a strong base of early literacy skills in preparation for later school success.

**Methods:** All fifteen classrooms that participated in the study were full-day programs serving children for the two years prior to their entry into kindergarten. These classrooms enrolled approximately 18-20 children per classroom with each class led by one lead teacher and one teacher's aide. The program took a three-tiered approach to



early literacy instruction. Tier 1 and Tier 2 were the teacher-implemented portions of the program; both involve professional development, weekly coaching sessions, and collaborative planning. Tier 3 involved individual children being tutored by specially-trained tutors from the University of Wisconsin. The program evaluation consisted of two components: progress monitoring and literacy environment observations. Child assessments were performed three times throughout the school year in September, January, and May using the PPVT-III, the PALS-PreK, and a story-telling measure created by the researchers. Progress monitoring required that the teachers document individual children's progress on a monthly basis. Literacy environment observations were performed using the ELLCO toolkit. In the first program year observations were performed in the beginning of the school year (September 2005) and at the end of the school year (May 2006).

**Findings:** In September 2005 the average ELLCO scores across the 15 classrooms were 47.18 (SD=7.47). The average ELLCO scores in May 2006 were 63.44 (SD=4.11). Researchers reported a significant increase in scores from the pre- to post-observation.

Hindman, A. H., and Wasik, B. (2008). Head Start teachers' beliefs about language and literacy instruction. *Early Childhood Research Quarterly*, 23, 479-492.

**Purpose:** This study examined various aspects of teachers' language and literacy beliefs. The goals of the study included determining whether or not a pre-existing measure of teacher literacy beliefs (measured by the Teacher Beliefs Questionnaire) accurately captured beliefs on alphabet knowledge and phonemic awareness as well as oral language and vocabulary, book reading, and writing. Researchers also wanted to determine whether teachers expressed agreement with pre-established "best practices" and what role teachers' background characteristics played in their expressed language and literacy beliefs. The study discusses how the specific content covered in the Teacher Beliefs Questionnaire corresponds with the aspects of quality focused on in the ELLCO. It does not report on data from use of the ELLCO, but rather discusses the constructs covered and not covered by the ELLCO.

**Methods:** Twenty-eight Head Start lead teachers were recruited as part of an Early Reading First professional development intervention in their centers. These teachers were surveyed on their literacy beliefs and practices using the Teacher Beliefs Questionnaire, comprised of 30 items addressing code-related skills, oral language/vocabulary, book reading, and writing.

**Findings:** In the discussion section of the article the authors begin to explore the ways in which teacher beliefs about language and literacy could be linked to classroom practice. The authors conclude that the ELLCO does not capture certain specific aspects of instruction considered to be important in the Teacher Beliefs Questionnaire, such as explicit and implicit teaching of letters, sounds, and word meanings. The authors also point out that the ELLCO does not include highly specific coding systems to capture precise amounts of instruction on various aspects of emergent literacy. Other studies have used such coding schemes to capture aspects of emergent literacy such as letter recognition, phonemic awareness, and decoding. The use of more highly specific coding schemes is recommended for the ELLCO.

Jackson, B., Larzelere, R., St. Clair, L., Corr, M., Fichter, C., and Egertson, H. (2006). The impact of HeadsUp! Reading on early childhood educators' literacy practices and preschool children's literacy skills. *Early Childhood Research Quarterly, 21*, 213-226.

**Purpose:** This paper examines the impact of teacher participation in a literacy-based professional development program called *HeadsUp! Reading* (HUR) on children's language and literacy scores. HUR is a 15-week satellite-broadcasted program aimed at improving teachers' language and literacy classroom practices for children from birth through age five. The researchers hypothesized that HUR would have significant positive impacts on teachers' language and literacy practices resulting in increases in children's language and literacy skills.

**Methods:** Study participants taught in a range of early childhood settings in the seven communities with the highest concentrations of poverty in Nebraska. The settings included Head Start centers, child care centers, federal Even Start Family Literacy programs, and state-funded pre-kindergarten programs. Teachers from 17 centers were assigned to the intervention group and a matched control group of 22 sites was identified based on ethnicity and language of the children and income of the families served. Participants in the intervention group (HUR group) participated in weekly HUR broadcasts each lasting approximately three hours spaced out over a fifteen-week period. Intervention group participants were given the opportunity to receive mentoring in addition to HUR broadcasts. The ELLCO and the ECERS-R were used at pre-intervention and post-intervention to assess each of the participating classrooms. Child assessments measuring vocabulary, language skills, reading skills, and teacher ratings of oral language were also performed pre- and post-intervention.

**Findings:** Researchers found that HUR group participants improved significantly more than control group participants on some ELLCO measures. Specifically, the HUR-only

group improved significantly more than the control group on the Literacy Environment Checklist. The HUR plus mentoring group improved significantly more than the control group on the Literacy Activities Rating Scale (LARS) total score and on the Writing subscale of the LARS. Additional analyses found that the Literacy Environment Checklist and the Literacy Activities Rating Scale significantly predicted an increase in Woodcock-Munoz Language scores. In addition, the Language, Literacy, and Curriculum subscale significantly predicted TERA reading quotient scores.

---

## Emerging Work

Bryant, D. (under review). Observational measures of quality in center-based early care and education programs.

**Purpose:** This paper reviews and compares multiple commonly used child care quality assessments. The predictive relationships between measures and child outcomes are reported upon. Criteria for selecting appropriate measures are also shared.

**Methods:** A systematic review and synthesis of existing literature was used to meet the paper's goals.

**Findings:** Multiple criteria for selecting a measure were offered. These criteria included: alignment with the study purpose (age of children targeted, whether the constructs of interest are measured), sound psychometric properties (such as validity and reliability of instrument), logistical and cost considerations (cost of training and administration, how easy it is for observers to be trained and remain reliable), and how the measure functions (is it sensitive to changes resulting from professional development, does it relate positively to desirable child outcomes). The choice between the use of empirically driven factors versus conceptually driven subscales to summarize data was related to the purpose of measurement, with the former being most commonly used for research purposes and the latter being most commonly used for program improvement. Multiple measures (ECERS-R, ECERS-E, ELLCO, ORCE, and Snapshot) were identified as sources for capturing cross-culturally appropriate care. Likewise, multiple measures (CIS, CLASS, ECCOM, ECERS-R, ORCE, PQA, Profile, and Snapshot) were related to program/staff characteristics. Finally, each of the measures reviewed (CLASS, ECERS-R, ECERS-E, ELLCO, ORCE, PQA, and Profile) were found to be predictive of desirable academic and/or social outcomes.

Burchinal, M., Kainz, K., and Cai, Y. (under review). How well are our measures of quality predicting to child outcomes: A meta-analysis and coordinated analysis of data from large scale studies of early childhood settings.

**Purpose:** This study uses meta-analysis and secondary data analysis to examine the relationship between measures of child care quality and child outcomes.

**Methods:** Two methods were used in this paper: meta-analysis and secondary data analysis. The meta-analysis included 19 peer-reviewed studies that measured the association between child care quality and child outcomes. Effect sizes from each study were then used to determine both the average effect size associated with various child outcomes (overall, academic/cognitive, language, and social) by age and type of outcome. The secondary data analysis used data from five major studies that include at least 100 low-income children in at least 50 classrooms (NICHD Study of Early Child Care and Youth Development; Cost, Quality, and Outcomes Study; National Center for Early Development and Learning 11-state Pre-Kindergarten Evaluation; and Head Start FACES study). The relationship between multiple instruments for measuring child care quality (ECERS, CIS, CLASS, and ORCE) and child outcomes were then analyzed with multivariate methods.

**Findings:** Using both methods, modest, but significant relationships were found between measures of child care quality and academic, language, and social-emotional child outcomes. Global quality measures tended to have more modest associations than specific measures/subscales capturing features of the environment most closely related to the child outcomes of interest (for example, language and literacy stimulation in the environment and children's vocabulary development). Based on a review of findings from previous studies, the authors hypothesized that stronger relationships between measures of child care quality and child outcomes might be detected if child care quality measures were aligned more directly with child outcomes of interest and if the psychometric properties of such measures were improved through testing of larger item pools and the use of item response theory.

Thornburg, K. R., Mauzy, D., Mayfield, W., Scott, J. L., Sparks, A., Mumford, J., Foulkes, T., and Fuger, K. L. (under review). Data-driven decision making in preparation for large-scale QRS implementation.

**Purpose:** This article reviews the practicalities of implementing a Quality Rating System in Missouri.

**Methods:** This article is a compilation and integration of lessons learned while implementing the Missouri Quality Rating System.

**Findings:** The Missouri Quality Rating System used the ECERS-R to measure the broad learning environment. The results of a pilot study revealed the ECERS-R to be a limited tool in assessing intentional teaching and curriculum. Thus, Missouri added an assessment of “Intentional Teaching,” measured through the ECERS-E, to their Quality Rating System. A unique scoring mechanism was used in which the average distribution of scores, in combination with minimum scores thresholds, defined quality observation criteria for each of the Quality Rating System’s tiers. As the ECERS-E was designed for three- to five-year-olds, Missouri researchers also developed the Infant/Toddler Intentional Teaching Checklist and the School-Aged Intentional Teaching Checklist. A standard of measuring half of the classrooms at each site was agreed upon as this yielded an 86 percent match to results from assessing all classrooms at a site.