



Quality in Early Childhood Care and Education Settings: A Compendium of Measures Second Edition

Project Coordinators:

Tamara Halle
Jessica Vick Whittaker
Rachel Anderson

Authors of Quality Measures Profiles: Mirjam Neunning, Debra Weinstein, Tamara Halle, Laurie Martin, Kathryn Tout, Laura Wandner, Jessica Vick Whittaker, Heather See, Meagan McSwiggan, Megan Fletcher, Juli Sherman, Elizabeth Hair, and Mary Burkhauser

2010

This document was prepared under Contract # HHSP233200800445G with the Administration for Children and Families, U S Department of Health and Human Services.

Prepared for:
Ivelisse Martinez-Beck, Ph.D.
U.S. Dept. of Health and Human Services
Administration for Children and Families
Office of Planning, Research and Evaluation
370 L'Enfant Plaza Promenade, SW
Washington, DC 20447

Quality in Early Childhood Care and Education Settings: A Compendium of Measures Second Edition

March 2010

Prepared by Child Trends
4301 Connecticut Avenue, NW, Suite 100
Washington, DC 20008
www.childtrends.org

Project Coordinators

Tamara Halle
Jessica Vick Whittaker
Rachel Anderson



This document was prepared under Contract # HHSP233200800445G with the Administration for Children and Families, U S Department of Health and Human Services, under the direction of project officer Ivelisse Martinez-Beck. Tamara Halle, Jessica Vick Whittaker, and Rachel Anderson are researchers at Child Trends. The views represented in this document are those of the authors and do not reflect the opinions of the Office of Planning, Research and Evaluation of the Administration for Children and Families.

Authors of Quality Measures Profiles

Mirjam Neunning, Debra Weinstein, Tamara Halle, Laurie Martin, Kathryn Tout, Laura Wandner, Jessica Vick Whittaker, Heather See, Meagan McSwiggan, Megan Fletcher, Juli Sherman, Elizabeth Hair, and Mary Burkhauser.

Suggested Citation:

Halle, T., Vick Whittaker, J. E., & Anderson, R. (2010). *Quality in Early Childhood Care and Education Settings: A Compendium of Measures, Second Edition*. Washington, DC: Child Trends. 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.

Please contact Tamara Halle at 4301 Connecticut Ave., NW, Suite 350, Washington, DC 20008, T: 202.572.6034, F: 202.362.5533, thalle@childtrends.org for more information about this publication.

Acknowledgements

The authors would like to thank the Office of Planning, Research and Evaluation of the Administration for Children and Families, U.S. Department of Health and Human Services for supporting this work. In particular, we are grateful to Dr. Ivelisse Martinez-Beck for identifying the need for this work and supporting the development of this document through her invaluable feedback and suggestions.

We would like to acknowledge the developers of the measures included in this compendium, and are especially grateful to those who took the time to review and provide feedback on the profiles of their measures. This helped to ensure that we have included the most updated information on the measures at this time.

The authors would also like to thank Dr. Martha Zaslow, Dr. Kathryn Tout, and Dr. Nicole Forry who helped conceptualize this project, provided input throughout its development, reviewed this document in its draft form, and provided feedback throughout the production of this document.

Finally, we would like to recognize and refer readers to three other measures compendia which may add supplementary information to that which is provided in this compendium.

- *Early Childhood Measures Profiles Compendium* (2004)

<http://aspe.hhs.gov/hsp/ECMeasures04/report.pdf>

This Early Childhood Measures Profiles Compendium, created for the SEED (Science and the Ecology of Early Development) Consortium of federal agencies, includes information on child outcome measures in the following domains: approaches to learning, general cognitive skills, language, literacy, math, ongoing observations, social-emotional functioning, and Early Head Start. The format and topic headings for measures profiles used in the current compendium were modeled after the SEED compendium of Early Childhood Measures.

- *Measuring Youth Program Quality: A Guide to Assessment Tools, Second Edition* (2009)

<http://www.forumforyouthinvestment.org/files/MeasuringYouthProgramQuality2ndEd.pdf>

This document, developed by The Forum for Youth Investment, is an excellent source for readers seeking information about measures that assess after-school and youth program quality. The authors write, "This guide was designed to compare the purpose, structure, content and technical properties of several youth program quality assessment tools" (Yohalem, N. and Wilson-Ahlstrom, A. with Fischer, S. and Shinn, M., 2009).

In order to avoid redundancies between measures compendia, we generally do not include profiles for quality measures for school-age settings in the current

compendium. An exception is the profile of the School Age Care Environment Rating Scale (SACERS; Harms, Vineberg, Jacobs, & Romano, 1996), which was included to provide a complete set of profiles for the Environmental Ratings Scales.

We are grateful to Nicole Yohalem, Alicia Wilson-Ahlstrom, Sean Fischer, and Marybeth Shinn for permitting us to reproduce portions of the Technical Glossary within *Measuring Youth Program Quality: A Guide to Assessment Tools* and to adopt the format for two summary tables from that compendium for use in this compendium.

- *Resources for Measuring Service and Outcomes in Head Start Programs Serving Infants and Toddlers* (2003)

http://www.acf.hhs.gov/programs/opre/ehs/perf_measures/reports/resources_measuring/resources_for_measuring.pdf

This document was developed by Mathematica Policy Research, Inc. for the Office of Planning, Research and Evaluation to help Early Head Start programs develop a performance measurement plan. It includes profiles of measures with detailed descriptions, scoring information, psychometrics, and information on training.

References

- Berry, D. J., Bridges, L. J., & Zaslow, M. J. (2004). *Early Childhood Measures Profiles*. Washington, D.C.: Prepared by Child Trends for The SEED Consortium of federal agencies which include the Office of the Assistant Secretary for Planning and Evaluation of the U.S. Department of Health and Human Services, the National Institute of Child Health and Human Development, and the Office of Planning, Research and Evaluation of the Administration for Children and Families of the U.S. Department of Health and Human Services.
- Kisker, E. E., Boller, K., Nagatoshi, C., Sciarrino, C., Jethwani, V., Zavitsky, T., Ford, M., & Love, J. M. (2003). *Resources for Measuring Services and Outcomes in Head Start Programs Serving Infants and Toddlers*. Washington, D.C.: Prepared by Mathematica Policy Research, Inc. for the Office of Planning, Research and Evaluation of the Administration for Children and Families of the U.S. Department of Health and Human Services.
- Yohalem, N. & Wilson-Ahlstrom, A, with Fischer, S., & Shinn, M. (2009, January). *Measuring Youth Program Quality: A Guide to Assessment Tools, Second Edition*. Washington, D.C.: The Forum for Youth Investment, Impact Strategies.

Table of Contents

Introduction	8
Cross - Cutting Comparisons	12
Target Age and Purpose	13
Methodology.....	18
Child Development Domains	23
Structure, Administration, and Staff Domains	29
Training and Administration	35
Technical Glossary.....	42
Profiles	44
Assessment Profile for Early Childhood Programs (APECP)	44
Assessment Profile for Family Child Care Homes (APFCCH)	44
Assessment of Practices in Early Elementary Classrooms (APEEC)	52
Business Administration Scale for Family Child Care (BAS)	57
Classroom Assessment of Supports for Emergent Bilingual Acquisition (CASEBA)	62
The Child Care Assessment Tool for Relatives (CCAT-R)	65
The Child Care HOME Inventories (CC-HOME)	71
Child Caregiver Interaction Scale (CCIS)	77
The Child-Caregiver Observation System (C-COS)	82
Child Development Program Evaluation Scale (CDPES)	88
Child/Home Early Language & Literacy Observation (CHELLO)	93
Arnett Caregiver Interaction Scale (CIS)	99
Classroom Assessment Scoring System (CLASS)	102
Classroom Assessment Scoring System: Toddler Version (CLASS Toddler).....	104
Classroom Language and Literacy Environment Observation (CLEO)	118
The Classroom Observation of Early Mathematics Environment and Teaching (COEMET)	124
Caregiver Observation Form and Scale (COFAS)	129
Classroom Practices Inventory (CPI)	132
The Emergent Academic Snapshot (Snapshot)	136
Early Childhood Classroom Observation Measure (ECCOM)	140
The Early Childhood Environment Rating Scale – Extension (ECERS-E)	145
Early Childhood Environment Rating Scale – Revised Edition (ECERS-R).....	150

Early Language & Literacy Classroom Observation (ELLCO)	158
Early Language & Literacy Classroom Observation Pre-K Tool (ELLCO Pre-K)	164
Early Language & Literacy Classroom Observation: Addendum for English Language Learners (ELLCO-ELL)	168
Early Literacy Observation Tool (E-LOT)	172
Emlen Scales: A Packet of Scales for Measuring the Quality of Child Care From a Parent's Point of View	177
Environment and Policy Assessment and Observation (EPAO)	187
Family Child Care Environment Rating Scale-Revised Edition (FCCERS-R).....	192
The Individualized Classroom Assessment Scoring System (inCLASS)	198
Infant/Toddler Environment Rating Scale Revised Edition (ITERS-R)	205
Language Interaction Snapshot (LISn)	212
Observation Measures of Language and Literacy (OMLIT)	217
Observational Record of the Caregiving Environment (ORCE)	226
Program Administration Scale (PAS)	240
Program for Infant/Toddler Care Program Assessment Rating Scale (PITC PARS).....	246
The Preschool Classroom Implementation Rating Scale (PCI)	255
Preschool Mental Health Climate Scale (PMHCS)*	259
Preschool Program Quality Assessment, 2 nd Edition (PQA)	262
Preschool Rating Instrument for Science and Math (PRISM)	268
Quality of Early Childhood Care Settings: Caregiver Rating Scale (QUEST)	272
Ramey and Ramey Observation of Learning Essentials (ROLE)	275
Ready School Assessment (RSA)	279
School – Age Care Environment Rating Scale (SACERS)*	284
Supports for Early Literacy Assessment (SELA)	290
Supports for Social-Emotional Growth Assessment (SSEGA)	295
Teacher Behavior Rating Scale (TBRS)	298
Teacher Instructional Engagement Scale (TIES)	306
Teacher Knowledge Assessment (TKA)	309
Teaching Pyramid Observation Tool for Preschool Classrooms (TPOT)*	312

Note. * Indicates that the measure has not been reviewed by the developer. The remaining profiles have been reviewed by the developer and updated according to his/her feedback.

Introduction

Quality measures were originally developed for research aimed at describing the settings in which children spend time and identifying the characteristics of these environments that contribute to children's development. They were also developed to guide improvements in practice. Increasingly, however, measures of quality are being used for further purposes. In particular, they are being used to guide components of state policies. For example, many states are developing Quality Initiatives and employing measures originally created for research or for guiding improvement in practice for the new purpose of assigning quality ratings to early care and education settings. States are also using these measures to monitor change in quality over time.

The *Quality in Early Childhood Care and Education Settings: A Compendium of Measures, Second Edition* was compiled by Child Trends for the Office of Planning, Research and Evaluation of the Administration for Children and Families, U.S. Department of Health and Human Services, to provide a consistent framework with which to review the existing measures of the quality of early care and education settings. The aim is to provide uniform information about quality measures. It is hoped that such information will be useful to researchers and practitioners, and help to inform the measurement of quality for policy-related purposes.

Criteria for Inclusion

The measures included in this compendium are in various stages of development. Some are new measures that have recently been validated, while

others have been in existence for over twenty years and have gone through several rounds of revision. Some of the included measures are widely known and recognized, and others are quickly gaining popularity.

The criteria for inclusion in this measures compendium were as follows:

- The measure is used in early care and education settings to assess the quality of the setting.
- Psychometric information about the measure is available (measures with forthcoming psychometric information are noted as "under development").
- The measure can be obtained for use.
- Where possible, the most current version of a measure is profiled; minor variations on a measure are not included.

Developers of the measures were contacted and asked to review the summaries for accuracy and completeness. Profiles were updated and revised based on input received from the developers. Three profiles are still under review; these profiles are identified as "under review" in the page header of the profiles, and have an asterisk next to their title in the Table of Contents.

We view this compendium as a document which will require updating on a periodic basis. Indeed, this second edition of the compendium, compared to the first which was released in 2007, includes 18 new measures profiles and 10 revised profiles. We list measures currently under development, for which

psychometric information is not yet available, below. We have included profiles for these measures (denoted with a hammer next to the title of the measure indicating that it is under development) and anticipate that additional psychometric information will be added to the profiles in a future version of this compendium.

Contents of the Compendium

The following information is included for each measure within this compendium:

- **Background Information**
 - Author and Publisher of the measure
 - Purpose of the Measure
 - Population Measure Developed With

- Age Range/Setting Intended For
- Ways in which Measure Addresses Diversity
- Key Constructs & Scoring of Measure
- Comments

- **Administration of Measure**
 - Who Administers Measure/Training Required
 - Setting
 - Time Needed and Cost

- **Functioning of Measure**
 - Reliability Information
 - Validity Information
 - Comments

Measures under Development*

- **Classroom Assessment of Supports for Emergent Bilingual Acquisition (CASEBA)**
- **Classroom Code for Interactive Recording of Children’s Learning Environments (Classroom CIRCLE)**
- **Early Language & Literacy Classroom Observation (ELLCO) Pre-K Tool**
- **Early Language & Literacy Classroom Observation: Addendum for English Language Learners (ELLCO: Addendum for ELL)**
- **Family Child Care Environment Rating Scale-Revised Edition (FCCERS-R)**
- **Preschool Mental Health Climate Scale (PMHCS)**
- **Preschool Rating Instrument for Science and Math (PRISM)**
- **Quality of Early Childhood Care Settings: Caregiver Rating Scale (QUEST)**
- **Ramey and Ramey Observation of Learning Essentials (ROLE)**
- **Teacher Knowledge Assessment (TKA)**
- **Teaching Pyramid Observation Tool for Preschool Classrooms (TPOT)**



*Measures under development are instruments for which extensive psychometric information is not yet available.

Updated Contents

In this Second Edition of the Compendium, we have included all of the measures included in the First Edition, along with 17 new measures. Since the release of the original Compendium, many of the measures developers have continued work on their assessment tools (e.g., validating with new samples, publishing new versions, updating scoring criteria). We provided all of the measures developers the opportunity to update their profiles for inclusion in this version of the Compendium. Ten developers have updated the profiles for their assessment tools. Additionally, there

has been great progress in the development of new assessment tools to assess the quality of early care and education settings. The 17 new tools represent the latest additions. As we continue to update the compendium, we will include new assessment tools, and provide developers with the opportunity to add the most up-to-date information on their assessment profiles. Please find below a list of tools that are new to this Second Edition of the Compendium, as well as a list of the measures that have been updated since the original version.

Assessment Tools Added to the Second Edition of the Compendium

- Business Administration Scale for Family Child Care (BAS)
- Classroom Assessment of Supports for Emergent Bilingual Acquisition (CASEBA)
- Classroom CIRCLE: Classroom Code for Interactive Recording of Children’s Learning Environments (CIRCLE)
- Classroom Assessment Scoring System: Toddler Version (CLASS Toddler)
- Classroom Observation of Early Mathematics Environment and Teaching (COEMET)
- Early Language & Literacy Classroom Observation – Addendum for English Language Learners (ELLCO – ELL)
- Early Language and Literacy Classroom Observation Pre-Kindergarten (ELLCO – Pre-K)
- Individualized Classroom Assessment Scoring System (inCLASS)
- Language Interaction Snapshot (LISn)
- Program for Infant/Toddler Care Program Assessment Rating Scale (PITC-PARS)
- Preschool Mental Health Climate Scale (PMHCS)
- Preschool Rating Instrument for Science and Math (PRISM)
- Teacher Behavior Rating Scale (TBRS)
- Instructional Engagement Scale (TIES)
- Teacher Knowledge Assessment (TKA)
- Teaching Pyramid Observation Tool for Preschool Classrooms (TPOT)
- Toddler Classroom Assessment Scoring System (Toddler CLASS)

Assessment Tools That Have Been Updated



- Assessment of Practices in Early Elementary Classrooms (APEEC)
- Child Care Assessment Tool for Relatives (CCATR)
- Child-Caregiver Observation System (C-COS)
- Child/Home Early Language & Literacy Observation (CHELLO)
- Classroom Assessment Scoring System (CLASS)
- Early Childhood Environmental Rating Scale – Revised Edition (ECERS-R)
- Environment and Policy Assessment and Observation (EPAO)
- Family Child Care Environment Rating Scale – Revised Edition (FCCERS-R)
- Program Administration Scale (PAS)
- Ready School Assessment (RSA)

Cross - Cutting Comparisons



The profiles of individual measures provide information about the technical properties of each. However, the following cross-cutting tables highlight the similarities and differences among the measures in several key areas. These tables are intended to aid users of the compendium in comparing various

features of the measures. Decisions about inclusion of measures in categories represented in the tables were based on a review of the measures themselves, manuals, relevant websites, and empirical studies in which the measures were used.

Table 1: Target Age and Purpose

The measures in this compendium were designed for a variety of purposes including program improvement, monitoring/accreditation, and research/evaluation. Many of them were designed to serve more than one purpose. The measures were developed to assess quality in diverse settings including home-based and center-based programs. All of the measures assess quality in early childhood care and education settings, but vary in the range of ages for which each measure is customized.

	Program Target Age & Setting	Primary Purposes		
		<i>Improvement</i>	<i>Monitoring/ Accreditation</i>	<i>Research/ Evaluation</i>
	<i>Ages/Setting Served</i>			
Assessment Profile for Early Childhood Programs (APECP)	Infant, toddler, pre-school and school-age programs	✓		✓
Assessment of Practices in Early Elementary Classrooms (APEEC)	K-3 general education classrooms		✓	✓
Assessment Profile for Family Child Care Homes (APFCCH)	Family child care homes for infants through school-age children	✓		✓
Business Administration Scale (BAS)	Family child care programs serving children of all ages	✓	✓	✓
Classroom Assessment of Supports for Emergent Bilingual Acquisition (CASEBA)	Pre-school classrooms where children are English Language Learners			✓
The Child Care Assessment Tool for Relatives (CCAT-R)	Family, friend, and neighbor care settings for children under age 6			✓
Child Care Home Inventories (CC-HOME)	Non-parental child care arrangements in home-like settings for children from infancy – age 6		✓	✓
Child Caregiver Interaction Scale (CCIS)	Children from infancy through school age in home and center based settings	✓	✓	✓
Child-Caregiver Observation System (C-COS)	1 to 5 year old children in all types of child care settings			✓

	Program Target Age & Setting	Primary Purposes		
		<i>Improvement</i>	<i>Monitoring/ Accreditation</i>	<i>Research/ Evaluation</i>
	<i>Ages/Setting Served</i>			
Child Development Program Evaluation Scale (CDPES)	Infant, toddler, pre-school, and school-age child care settings		✓	✓
Child/Home Early Language & Literacy Observation (CHELLO)	Mixed-age, home-based care settings	✓		✓
Caregiver (Adult) Interaction Scale (CIS)	Early childhood classrooms or family child care homes			✓
Classroom Assessment Scoring System (CLASS)	Two versions of the CLASS are available: a pre-school classroom version and a K-3 classroom version	✓		✓
Classroom Assessment Scoring System: Toddler Version (CLASS Toddler)	Center-based child care classrooms serving children 15 to 26 months	✓		✓
Classroom CIRCLE	Pre-school programs serving children 6 to 72 months			✓
Classroom Language and Literacy Environment Observation (CLEO)	Formal child care or Head Start settings serving children ages 3-5			✓
Classroom Observation of Early Mathematics Environment and Teaching (COEMET)	Early childhood settings for toddlers -- 2 nd grade.			✓
Child Observation Form and Scale (COFAS)	Infant – 12 year-old classroom settings		✓	✓
Classroom Practices Inventory (CPI)	Early childhood programs for 4 and 5-year olds. This measure was adapted for use in kindergarten-primary programs		✓	✓
The Emergent Academic Snapshot (EAS)	Pre-school – kindergarten aged children in child care, pre-school, or kindergarten settings			✓

	Program Target Age & Setting	Primary Purposes		
	<i>Ages/Setting Served</i>	<i>Improvement</i>	<i>Monitoring/ Accreditation</i>	<i>Research/ Evaluation</i>
Early Childhood Classroom Observation Measure (ECCOM)	Classrooms serving children ages 4 - 7	✓		✓
Early Childhood Environment Rating Scale Extension (ECERS-E)	Early childhood classrooms serving 3 – 5 year olds	✓		✓
Early Childhood Environment Rating Scale – Revised (ECERS-R)	Early childhood classrooms serving 2 ½ – 5 year olds	✓	✓	✓
Early Language & Literacy Classroom Observation (ELLCO)	Pre-K to 3 rd grade classrooms	✓		✓
Early Language & Literacy Classroom Observation Pre-K Tool (ELLCO – Pre-K)	Center-based classrooms for 3- to 5-year-old children	✓		✓
Early Language & Literacy Classroom Observation: Addendum for English Language Learners (ELLCO: Addendum for ELL)	Pre-K -- 3 rd grade classrooms	✓		✓
Early Literacy Observation Tool (E-LOT)	Pre-K and Kindergarten classrooms (There is also a version for elementary classrooms)	✓	✓	✓
Emlen Scales	Children of all ages in any type of child care arrangement			✓
Environment and Policy Assessment and Observation (EPAO)	Child care centers for pre-school aged children		✓	
Family Child Care Environment Rating Scale – Revised Edition (FCCRS-R)	Family child care home settings serving children birth – elementary school		✓	✓
Individualized Classroom Assessment Scoring System (inCLASS)	Pre-school and kindergarten programs for 3- to 5-year olds in all settings	✓		✓
Infant and Toddler Environment Rating Scale – Revised (ITERS-R)	Classrooms serving infants from birth -- 30 months	✓		✓

	Program Target Age & Setting	Primary Purposes		
	<i>Ages/Setting Served</i>	<i>Improvement</i>	<i>Monitoring/ Accreditation</i>	<i>Research/ Evaluation</i>
Language Interaction Snapshot (LISn)	Family and center-based classrooms serving 3- and 4-year olds			✓
Observation Measures of Language and Literacy Instruction (OMLIT)	Early childhood classrooms; The Snapshot and Read Aloud Profile have also been used in family child care homes			✓
Observational Record of the Caregiving Environment (ORCE)	Primary child care settings – measures available at ages 6,15,24,36, and 54 months			✓
Program Administration Scale (PAS)	Center-based or public school-based early care and education programs	✓	✓	✓
Preschool Classroom Implementation Rating Scale (PCI)	Pre-school or kindergarten classrooms in public or private schools, day care centers, Head Start, or church programs serving children ages 3-6		✓	✓
Program for Infant/Toddler Care Program Assessment Rating Scale (PITC PARS)	Programs that serve children ages 0-3 outside the home	✓	✓	✓
Preschool Mental Health Climate Scale (PMHCS)	Pre-school aged children in Head Start or other pre-school programs	✓		✓
Preschool Program Quality Assessment Instrument (PQA)	All center-based pre-school settings, regardless of whether center is using High Scope	✓	✓	✓
Preschool Rating Instrument for Science and Math (PRISM)	Pre-school classrooms			✓
Quality of Early Childhood Care Settings: Caregiver Rating Scale (QUEST)	Variety of settings from informal care to formal center-based care of children 0-5 years of age			✓

	Program Target Age & Setting	Primary Purposes		
		<i>Improvement</i>	<i>Monitoring/ Accreditation</i>	<i>Research/ Evaluation</i>
Ramey's Observation of the Learning Environment (ROLE)	Pre-K classrooms serving 3- and 4-year old children			✓
Ready School Assessment (RSA)	Elementary schools with an emphasis on pre-K – 2 nd grade classrooms	✓		
School Age Care Environment Rating Scale (SACERS)	School-aged care settings serving children ages 5 -- 12 years old	✓	✓	✓
Supports for Early Literacy Assessment (SELA)	Center-based pre-school settings serving children ages 3 -- 5			✓
Supports for Social-Emotional Growth Assessment (SSEGA)	Children ages 3 -- 5 in center-based pre-school settings			✓
Teacher Behavior Rating Scale (TBRS)	Teachers and caregivers of children ages 3 -- 5 in a variety of settings	✓	✓	✓
Teacher Instructional Engagement Scale	Pre-school classrooms serving low-income children			✓
Teacher Knowledge Assessment (TKA)	Early education teachers who are new to the field or who have completed professional development in language and literacy	✓		✓
Teaching Pyramid Observation Tool for Preschool Classrooms (TPOT)	Pre-school classrooms			✓

The format for this table is from *Measuring Youth Program Quality: A Guide to Assessment Tools* (Yohalem and Wilson-Ahlstrom with Fischer and Shinn, 2009) and was used with permission from Nicole Yohalem. The content of this table is specific to *Quality in Early Childhood Care and Education Settings: A Compendium of Measures, Second Edition* (Halle, Vick Whittaker, & Anderson, 2010) and was compiled by the authors.

Table 2: Methodology

Most of the measures included in this compendium are intended for external observers, although some of the tools may be used by center directors and teachers for program monitoring and improvement. Many developers stress the importance of training users before implementing the measures. Observation is the primary data collection method employed, although some instruments supplement the observations with interviews, questionnaires, and document review.

	Target Users		Data Collection Methods			
	<i>Program* Staff</i>	<i>External Observers</i>	<i>Observation</i>	<i>Interview</i>	<i>Questionnaire</i>	<i>Document Review</i>
Assessment Profile for Early Childhood Programs (APECP)	✓	✓	✓	✓		✓
Assessment of Practices in Early Elementary Classrooms (APEEC)		✓	✓	✓		
Assessment Profile for Family Child Care Homes (APFCCH)		✓	✓	✓		✓
Business Administration Scale (BAS)	✓	✓		✓		✓
Classroom Assessment of Supports for Emergent Bilingual Acquisition (CASEBA)		✓	✓	✓		
The Child Care Assessment Tool for Relatives (CCAT-R)		✓	✓	✓		
Child Care Home Inventories (CC-HOME)		✓	✓			
Child Caregiver Interaction Scale (CCIS)		✓	✓			
Child-Caregiver Observation System (C-COS)		✓	✓			
Child Development Program Evaluation Scale (CDPES)	✓	✓	✓	✓		✓
Child/Home Early Language & Literacy Observation (CHELLO)		✓	✓	✓		
Caregiver (Adult) Interaction Scale (CIS)		✓	✓			

	Target Users		Data Collection Methods			
	<i>Program* Staff</i>	<i>External Observers</i>	<i>Observation</i>	<i>Interview</i>	<i>Questionnaire</i>	<i>Document Review</i>
Classroom Assessment Scoring System (CLASS)		✓	✓			
Classroom Assessment Scoring System: Toddler Version (CLASS Toddler)		✓	✓			
Classroom CIRCLE		✓	✓			
Classroom Language and Literacy Environment Observation (CLEO)		✓	✓			
Classroom Observation of Early Mathematics Environment and Teaching (COEMET)		✓	✓			
Child Observation Form and Scale (COFAS)	✓	✓	✓			
Classroom Practices Inventory (CPI)		✓	✓			
The Emergent Academic Snapshot (EAS)		✓	✓			
Early Childhood Classroom Observation Measure (ECCOM)	✓	✓	✓			
Early Childhood Environment Rating Scale Extension (ECERS-E)	✓	✓	✓	✓		
Early Childhood Environment Rating Scale – Revised (ECERS-R)	✓	✓	✓	✓		
Early Language and Literacy Classroom Observation (ELLCO)	✓	✓	✓	✓	✓	
Early Language & Literacy Classroom Observation Pre-K Tool (ELLCO – Pre-K)	✓	✓	✓	✓		
Early Language & Literacy Classroom Observation: Addendum for English						

	Target Users		Data Collection Methods			
	<i>Program* Staff</i>	<i>External Observers</i>	<i>Observation</i>	<i>Interview</i>	<i>Questionnaire</i>	<i>Document Review</i>
Language Learners (ELLCO: Addendum for ELL)	✓	✓	✓	✓	✓	
Early Literacy Observation Tool (ELOT)	✓	✓	✓			
Emlen Scales	✓ (Parents)				✓	
Environment and Policy Assessment and Observation (EPAO)		✓	✓			✓
Family Child Care Environment Rating Scale – Revised Edition (FCCRS-R)	✓	✓	✓	✓		
Individualized Classroom Assessment Scoring System (inCLASS)	✓ (Teachers and Parents)	✓	✓		✓	
Infant and Toddler Environment Rating Scale – Revised (ITERS-R)	✓	✓	✓	✓		
Language Interaction Snapshot (LISn)		✓	✓			
Observation Measures of Language and Literacy Instruction (OMLIT)		✓	✓			
Observational Record of the Caregiving Environment (ORCE)		✓	✓			
Program Administration Scale (PAS)	✓	✓		✓		✓
Preschool Classroom Implementation Rating Scale (PCI)		✓	✓			
Program for Infant/Toddler Care Program Assessment Rating Scale (PITC)						

	Target Users		Data Collection Methods			
	<i>Program* Staff</i>	<i>External Observers</i>	<i>Observation</i>	<i>Interview</i>	<i>Questionnaire</i>	<i>Document Review</i>
PARS)		✓	✓	✓		✓
Preschool Mental Health Climate Scale (PMHCS)		✓	✓			
Preschool Program Quality Assessment Instrument (PQA)	✓	✓	✓	✓		
Preschool Rating Instrument for Science and Math (PRISM)		✓	✓			
Quality of Early Childhood Care Settings: Caregiver Rating Scale (QUEST)		✓	✓			
Ramey's Observation of the Learning Environment (ROLE)		✓	✓			
Ready School Assessment (RSA)	✓ (Parents and Principals)		✓		✓	✓
School Age Care Environment Rating Scale (SACERS)	✓	✓	✓	✓		
Supports for Early Literacy Assessment (SELA)		✓	✓	✓		
Supports for Social-Emotional Growth Assessment (SSEGA)		✓	✓	✓		
Teacher Behavior Rating Scale (TBRs)	✓	✓	✓			
Teacher Instructional Engagement Scale		✓	✓			
Teacher Knowledge Assessment (TKA)	✓				✓	
Teaching Pyramid Observation Tool for Preschool Classrooms (TPOT)		✓	✓	✓		

* Program staff refers to teachers, child care providers, and center directors. Other program staff is noted in parentheses.

The format for this table is from *Measuring Youth Program Quality: A Guide to Assessment Tools* (Yohalem and Wilson-Ahlstrom with Fischer and Shinn, 2009) and was used with permission from Nicole Yohalem. The content of this table is specific to *Quality in Early Childhood Care and Education Settings: A Compendium of Measures, Second Edition* (Halle, Vick Whittaker, & Anderson, 2010) and was compiled by the authors.

Table 3: Child Domains Covered – Quality Features Related to Child Development

Many of the measures included in this compendium assess features of quality within early care and education environments that are believed to support specific domains of child development and skills associated with school readiness and later school success. These domains include language development, literacy, math, science, creative arts, and general cognition, as well as social and emotional development, approaches to learning, and health/physical development. Many developers emphasize supports for either social-emotional or language and cognitive skills with their measures. However, some measures attempt to cover quite a broad range of supports for child development.

It should be noted that the check marks in this table do not connote breadth or depth of coverage of a particular developmental domain. Rather, a check mark indicates whether the measure addresses a certain domain *at all*. For example, one measure could have a single item that addresses supports for health/physical development, and another measure could be entirely focused on supports for health/physical development, and both measures would receive a check mark for health/physical development in this table. The general cognition category is mutually exclusive from subject categories such as math and science. For example, a measure that addresses math does not automatically also receive a check for general cognition; general cognition is only checked where a measure lists cognitive development or activities that are not related to a particular subject.

	Domains – Child Development								
	LA = Language Development, LI = Literacy, M = Math, S = Science, CA = Creative Arts, GC = General Cognition, SED = Social and Emotional Development, AL = Approaches to Learning, HP = Health/Physical Development								
	LA	LI	M	S	CA	GC	SED	AL	HP
Assessment Profile for Early Childhood Programs (APECP)						✓	✓	✓	✓
Assessment of Practices in Early Elementary Classrooms (APEEC)							✓	✓	✓
Assessment Profile for Family Child Care Homes (APFCCH)						✓	✓		✓
Business Administration Scale (BAS)									

	Domains – Child Development								
	LA = Language Development, LI = Literacy, M = Math, S = Science, CA = Creative Arts, GC = General Cognition, SED = Social and Emotional Development, AL = Approaches to Learning, HP = Health/Physical Development								
	LA	LI	M	S	CA	GC	SED	AL	HP
Classroom Assessment of Supports for Emergent Bilingual Acquisition (CASEBA)	✓	✓					✓	✓	
The Child Care Assessment Tool for Relatives (CCAT-R)	✓					✓	✓		✓
Child Care Home Inventories (CC-HOME)	✓					✓			
Child Caregiver Interaction Scale (CCIS)	✓	✓				✓	✓	✓	✓
Child-Caregiver Observation System (C-COS)							✓	✓	
Child Development Program Evaluation Scale (CDPES)	✓				✓	✓	✓		✓
Child/Home Early Language & Literacy Observation (CHELLO)	✓	✓					✓		
Caregiver (Adult) Interaction Scale (CIS)							✓		
Classroom Assessment Scoring System (CLASS)	✓					✓	✓	✓	
Classroom Assessment Scoring System: Toddler Version (CLASS Toddler)	✓					✓	✓	✓	
Classroom CIRCLE	✓	✓			✓		✓	✓	
Classroom Language and Literacy Environment Observation (CLEO)	✓	✓					✓		

	Domains – Child Development								
	LA = Language Development, LI = Literacy, M = Math, S = Science, CA = Creative Arts, GC = General Cognition, SED = Social and Emotional Development, AL = Approaches to Learning, HP = Health/Physical Development								
	<i>LA</i>	<i>LI</i>	<i>M</i>	<i>S</i>	<i>CA</i>	<i>GC</i>	<i>SED</i>	<i>AL</i>	<i>HP</i>
Classroom Observation of Early Mathematics Environment and Teaching (COEMET)			✓						
Child Observation Form and Scale (COFAS)	✓					✓	✓	✓	
Classroom Practices Inventory (CPI)							✓	✓	
The Emergent Academic Snapshot (EAS)	✓						✓	✓	
Early Childhood Classroom Observation Measure (ECCOM)		✓	✓		✓	✓	✓	✓	
Early Childhood Environment Rating Scale Extension (ECERS-E)	✓	✓	✓	✓					
Early Childhood Environment Rating Scale – Revised (ECERS-R)	✓	✓	✓	✓	✓		✓	✓	✓
Early Language and Literacy Classroom Observation (ELLCO)	✓	✓							
Early Language & Literacy Classroom Observation Pre-K Tool (ELLCO – Pre-K)	✓	✓							
Early Language & Literacy Classroom Observation: Addendum for English Language Learners (ELLCO: Addendum for ELL)	✓	✓							
Early Literacy Observation Tool (ELOT)	✓	✓							

	Domains – Child Development								
	LA = Language Development, LI = Literacy, M = Math, S = Science, CA = Creative Arts, GC = General Cognition, SED = Social and Emotional Development, AL = Approaches to Learning, HP = Health/Physical Development								
	<i>LA</i>	<i>LI</i>	<i>M</i>	<i>S</i>	<i>CA</i>	<i>GC</i>	<i>SED</i>	<i>AL</i>	<i>HP</i>
Emlen Scales							✓		✓
Environment and Policy Assessment and Observation (EPAO)									✓
Family Child Care Environment Rating Scale – Revised Edition (FCCRS-R)	✓	✓	✓	✓	✓		✓		✓
Individualized Classroom Assessment Scoring System (inCLASS)							✓	✓	
Infant and Toddler Environment Rating Scale – Revised (ITERS-R)	✓						✓		✓
Language Interaction Snapshot (LISn)	✓								
Observation Measures of Language and Literacy Instruction (OMLIT)	✓	✓						✓	
Observational Record of the Caregiving Environment (ORCE)	✓	✓				✓	✓	✓	
Program Administration Scale (PAS)									
Preschool Classroom Implementation Rating Scale (PCI) **	✓	✓			✓		✓	✓	✓
Program for Infant/Toddler Care Program Assessment Rating Scale (PITC PARS)							✓		✓

	Domains – Child Development								
	LA = Language Development, LI = Literacy, M = Math, S = Science, CA = Creative Arts, GC = General Cognition, SED = Social and Emotional Development, AL = Approaches to Learning, HP = Health/Physical Development								
	<i>LA</i>	<i>LI</i>	<i>M</i>	<i>S</i>	<i>CA</i>	<i>GC</i>	<i>SED</i>	<i>AL</i>	<i>HP</i>
Preschool Mental Health Climate Scale (PMHCS)						✓	✓	✓	
Preschool Program Quality Assessment Instrument (PQA)							✓		✓
Preschool Rating Instrument for Science and Math (PRISM)			✓	✓		✓		✓	
Quality of Early Childhood Care Settings: Caregiver Rating Scale (QUEST)	✓	✓				✓	✓	✓	✓
Ramey's Observation of the Learning Environment (ROLE)	✓	✓	✓	✓	✓	✓	✓		
Ready School Assessment (RSA)									✓
School Age Care Environment Rating Scale (SACERS)	✓	✓	✓	✓	✓		✓	✓	✓
Supports for Early Literacy Assessment (SELA)	✓	✓							
Supports for Social-Emotional Growth Assessment (SSEGA)							✓		
Teacher Behavior Rating Scale (TBRS)	✓	✓	✓				✓	✓	✓
Teacher Instructional Engagement Scale	✓	✓						✓	
Teacher Knowledge Assessment (TKA)	✓	✓					✓		✓
Teaching Pyramid Observation Tool for Preschool Classrooms							✓	✓	

Domains – Child Development									
LA = Language Development, LI = Literacy, M = Math, S = Science, CA = Creative Arts, GC = General Cognition, SED = Social and Emotional Development, AL = Approaches to Learning, HP = Health/Physical Development									
	LA	LI	M	S	CA	GC	SED	AL	HP
(TPOT)									

The format and content of this table is specific to *Quality in Early Childhood Care and Education Settings: A Compendium of Measures, Second Edition* (Halle, Vick Whittaker, & Anderson, 2010) and was compiled by the authors.

Table 4: Domains Covered- Quality Features Related to Structure, Administration, and Staff

Many of the measures included in this compendium assess features of quality within early care and education environments that reflect administrative structures and practices as well as supports for staff development. There are several areas related to *structure*, including business practices, family involvement, activities/scheduling, classroom organization, and classroom materials. There are also several areas related to *administration*, including internal communications and leadership/management. Finally, quality features related to *monitoring/improvement* include professional development, assessment/monitoring of children, and program/staff assessments. Examples of each of these aspects of quality are offered below:

Structure

Business Practices: Program has a method for keeping business records (financial or programmatic) or has sound, consistent business practices and procedures.

Family Involvement: There are specific practices in place to ensure communication with and/or involvement of families. Family involvement is viewed as important.

Activities/Scheduling: Observers are instructed to look for specific activities (i.e. circle time, outdoors time) and/or general schedule planning and facilitation (i.e. that there is a schedule, schedule flows and has good transitions, there are procedures for hand-washing, snack time, etc).

Classroom Organization: Refers to the physical layout of the program (i.e. well-defined spaces for different activities, specific areas are present--dramatic play area, outdoor playground-or materials or facilities are in good condition).

Classroom Materials: Classroom has either specific materials (i.e. blocks, books), a variety of materials, and/or materials that are developmentally appropriate.

Administration

Internal Communication: Leadership communicates well with staff and/or staff communicate well with each other.

Leadership/Management: Program director or principal plays an active, positive role in the functioning of the program, there is strong internal leadership, or teachers manage the program well.

Monitoring/Improvement

Professional Development: Professional development is supported and/or specific professional development opportunities (internal or external) are made available to staff.

Assessments/Monitoring of Children: Program completes assessments or monitoring of children (can refer to specific published assessment tools or to program-specific techniques).

Program/Staff Assessments: Program completes assessments or monitoring of the program and/or the staff (can refer to specific published assessment tools or to program-specific techniques).

	Domains – Structure, Administration, Staff									
	Domains: BP=Business Practices, FI=Family Involvement, AS=Activities/Scheduling, CO=Classroom Organization, CM=Classroom Materials, IC=Internal Communication, LM=Leadership/Management PD=Professional Development, AMS=Assessments/Monitoring of Students, PSA=Program/Staff Assessments									
	<i>Structure</i>					<i>Administration</i>		<i>Monitoring and Improvement</i>		
	<i>BP</i>	<i>FI</i>	<i>AS</i>	<i>CO</i>	<i>CM</i>	<i>IC</i>	<i>LM</i>	<i>PD</i>	<i>AMS</i>	<i>PSA</i>
Assessment Profile for Early Childhood Programs (APECP)	✓	✓	✓	✓	✓		✓	✓	✓	
Assessment of Practices in Early Elementary Classrooms (APEEC)		✓	✓	✓	✓				✓	
Assessment Profile for Family Child Care Homes (APFCCH)		✓		✓	✓			✓		
Business Administration Scale (BAS)	✓	✓					✓	✓		
Classroom Assessment of Supports for Emergent Bilingual Acquisition (CASEBA)			✓	✓	✓				✓	
The Child Care Assessment Tool for Relatives (CCAT-R)		✓								
Child Care Home Inventories (CC-HOME)			✓	✓	✓					

Domains – Structure, Administration, Staff

Domains: BP=Business Practices, FI=Family Involvement, AS=Activities/Scheduling, CO=Classroom Organization, CM=Classroom Materials, IC=Internal Communication, LM=Leadership/Management PD=Professional Development, AMS=Assessments/Monitoring of Students, PSA=Program/Staff Assessments

	Structure					Administration		Monitoring and Improvement		
	<i>BP</i>	<i>FI</i>	<i>AS</i>	<i>CO</i>	<i>CM</i>	<i>IC</i>	<i>LM</i>	<i>PD</i>	<i>AMS</i>	<i>PSA</i>
Child Caregiver Interaction Scale (CCIS)		✓	✓		✓					
Child-Caregiver Observation System (C-COS)			✓							
Child Development Program Evaluation Scale (CDPES)		✓	✓		✓	✓		✓	✓	✓
Child/Home Early Language & Literacy Observation (CHELLO)		✓	✓	✓	✓				✓	
Caregiver (Adult) Interaction Scale (CIS)										
Classroom Assessment Scoring System (CLASS)			✓	✓	✓					
Classroom Assessment Scoring System: Toddler Version (CLASS Toddler)			✓							
Classroom CIRCLE			✓							
Classroom Language and Literacy Environment Observation (CLEO)			✓		✓					
Classroom Observation of Early Mathematics Environment and Teaching (COEMET)			✓		✓				✓	
Child Observation Form and Scale (COFAS)			✓							
Classroom Practices Inventory (CPI)			✓							

Domains – Structure, Administration, Staff

Domains: BP=Business Practices, FI=Family Involvement, AS=Activities/Scheduling, CO=Classroom Organization, CM=Classroom Materials, IC=Internal Communication, LM=Leadership/Management PD=Professional Development, AMS=Assessments/Monitoring of Students, PSA=Program/Staff Assessments

	Structure					Administration		Monitoring and Improvement		
	<i>BP</i>	<i>FI</i>	<i>AS</i>	<i>CO</i>	<i>CM</i>	<i>IC</i>	<i>LM</i>	<i>PD</i>	<i>AMS</i>	<i>PSA</i>
	The Emergent Academic Snapshot (EAS)			✓						
Early Childhood Classroom Observation Measure (ECCOM)			✓		✓					
Early Childhood Environment Rating Scale Extension (ECERS-E)			✓		✓					
Early Childhood Environment Rating Scale – Revised (ECERS-R)		✓	✓	✓	✓	✓	✓	✓		✓
Early Language and Literacy Classroom Observation (ELLCO)		✓		✓	✓					
Early Language & Literacy Classroom Observation Pre-K Tool (ELLCO – Pre-K)		✓		✓	✓					
Early Language & Literacy Classroom Observation: Addendum for English Language Learners (ELLCO: Addendum for ELL)		✓		✓	✓					
Early Literacy Observation Tool (ELOT)					✓				✓	
Emlen Scales		✓	✓		✓					
Environment and Policy Assessment and Observation (EPAO)			✓	✓	✓					

Domains – Structure, Administration, Staff

Domains: BP=Business Practices, FI=Family Involvement, AS=Activities/Scheduling, CO=Classroom Organization, CM=Classroom Materials, IC=Internal Communication, LM=Leadership/Management PD=Professional Development, AMS=Assessments/Monitoring of Students, PSA=Program/Staff Assessments

	Structure					Administration		Monitoring and Improvement		
	<i>BP</i>	<i>FI</i>	<i>AS</i>	<i>CO</i>	<i>CM</i>	<i>IC</i>	<i>LM</i>	<i>PD</i>	<i>AMS</i>	<i>PSA</i>
Family Child Care Environment Rating Scale – Revised Edition (FCCRS-R)		✓	✓	✓	✓		✓	✓		
Individualized Classroom Assessment Scoring System (inCLASS)										
Infant and Toddler Environment Rating Scale – Revised (ITERS-R)		✓	✓	✓	✓					
Language Interaction Snapshot (LISn)			✓							
Observation Measures of Language and Literacy Instruction (OMLIT)			✓	✓	✓					
Observational Record of the Caregiving Environment (ORCE)			✓	✓						
Program Administration Scale (PAS)	✓	✓		✓		✓	✓	✓	✓	✓
Preschool Classroom Implementation Rating Scale (PCI)			✓	✓	✓	✓			✓	
Program for Infant/Toddler Care Program Assessment Rating Scale (PITC PARS)		✓	✓		✓					
Preschool Mental Health Climate Scale (PMHCS)			✓			✓				
Preschool Program Quality Assessment Instrument (PQA)	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Domains – Structure, Administration, Staff

Domains: BP=Business Practices, FI=Family Involvement, AS=Activities/Scheduling, CO=Classroom Organization, CM=Classroom Materials, IC=Internal Communication, LM=Leadership/Management PD=Professional Development, AMS=Assessments/Monitoring of Students, PSA=Program/Staff Assessments

	Structure					Administration		Monitoring and Improvement		
	BP	FI	AS	CO	CM	IC	LM	PD	AMS	PSA
Preschool Rating Instrument for Science and Math (PRISM)			✓		✓					
Quality of Early Childhood Care Settings: Caregiver Rating Scale (QUEST)			✓	✓	✓					
Ramey's Observation of the Learning Environment (ROLE)			✓						✓	
Ready School Assessment (RSA)		✓	✓		✓	✓	✓	✓	✓	✓
School Age Care Environment Rating Scale (SACERS)		✓	✓	✓	✓	✓	✓	✓		✓
Supports for Early Literacy Assessment (SELA)		✓	✓	✓	✓					
Supports for Social-Emotional Growth Assessment (SSEGA)		✓	✓	✓					✓	
Teacher Behavior Rating Scale (TBRS)			✓	✓	✓		✓		✓	
Teacher Instructional Engagement Scale			✓							
Teacher Knowledge Assessment (TKA)		✓				✓	✓		✓	
Teaching Pyramid Observation Tool for Preschool Classrooms (TPOT)		✓	✓	✓	✓					

The format and content of this table is specific to *Quality in Early Childhood Care and Education Settings: A Compendium of Measures, Second Edition* (Halle, Vick Whittaker, & Anderson, 2010) and was compiled by the authors.

Table 5: Training and Administration

As quality measures become more widely used within statewide Quality Initiatives, considerations about training and administration costs have become more salient, especially if states feel they need to use more than one quality measure to obtain all of the pertinent information they seek. However, researchers and program administrators who use quality measures for research or guiding improvement in practice are also interested in training and administration costs. This cross-cutting table offers a summary of key information on the time and financial investments associated with training and administration of quality measures. Please note that these costs may change over time; it is best to contact the developer for updated information.

	Training		Administration	
	<i>Training Needed</i>	<i>Cost of Training</i>	<i>Time Needed for Administration</i>	<i>Cost of Measure</i>
Assessment Profile for Early Childhood Programs (APECP)	2 – 3 day training	Information Not Available	4 – 6 hours (for administrative component only) or 1 day (to observe 2 – 3 classrooms)	\$18 for assessment profiles for 3 classrooms, \$25 for technical manual
Assessment of Practices in Early Elementary Classrooms (APEEC)	Information Not Available	Information Not Available	1 day	\$13.95
Assessment Profile for Family Child Care Homes (APFCCH)	2 – 3 day training	Information Not Available	4 – 6 hours	\$18 for assessment profiles for 3 classrooms, \$25 for technical manual
Business Administration Scale (BAS)	2 day training for accountability, monitoring and research. No training required for self-assessment.	Contact Jill Bella, Director of Special Projects at (800) 443-5522, ext. 5059 or jill.bella@nl.edu.	Approximately 2 hours	\$19.95
Classroom Assessment of Supports for Emergent Bilingual Acquisition (CASEBA)	At least 4 days training	Varies; researchers should be trained by the developers	Full session for half-day program, from start of day until nap time for full-day program	Varies
The Child Care Assessment				

	Training		Administration	
	<i>Training Needed</i>	<i>Cost of Training</i>	<i>Time Needed for Administration</i>	<i>Cost of Measure</i>
Tool for Relatives (CCAT-R)	2.5 day training	\$1075 per person	2.5 - 3 hours	\$250 for scoring software program
Child Care Home Inventories (CC-HOME)	No formal training; ½ day of independent training plus instrument practice has been used	No Formal Training	Approximately 1 hour	\$30 for manual, \$25 for 50 forms
Child Caregiver Interaction Scale (CCIS)	1 day training plus 2 practice observations	Contact the author: B.E.Carl@iup.edu	3 hours	Contact the author: B.E.Carl@iup.edu
Child-Caregiver Observation System (C-COS)	1 day classroom training and 2 field observations	Contact Mathematica Policy Research, Inc. (609) 799-3535 www.mathematica-mpr.com	2 - 3 hours	\$250 for audio prompts and training materials
Child Development Program Evaluation Scale (CDPES)	Approximately 1 week (1 - 2 days training plus 2 - 3 days on-site training)	Information Not Available	1 day for up to 60 children, 2 days for 61-120 children, 3 days for more than 121 children	None
Child/Home Early Language & Literacy Observation (CHELLO)	1 day training	Contact Dr. Susan Neuman:sbneuman@umich.edu	1.5 - 2 hours	Contact Dr. Susan Neuman:sbneuman@umich.edu
Caregiver (Adult) Interaction Scale (CIS)	Must achieve .70 inter-rater reliability across two consecutive visits to be certified	No Formal Training	At least 45 minutes	None
Classroom Assessment Scoring System (CLASS)	2 day training	\$670 per person	Approximately 2 hours	\$49.50 for Pre-K Manual, \$25 for 10 scoring forms
Classroom Assessment Scoring System: Toddler Version (CLASS Toddler)	Information Not Available	Information Not Available	Minimum of 2 hours	Information Not Available

	Training		Administration	
	<i>Training Needed</i>	<i>Cost of Training</i>	<i>Time Needed for Administration</i>	<i>Cost of Measure</i>
Classroom CIRCLE	2 - 4 days of training	Contact Dr. Jane Atwater: janea@ku.edu	Approximately 1.5 hours	Contact Dr. Jane Atwater: janea@ku.edu
Classroom Language and Literacy Environment Observation (CLEO)	1 day training plus 6 – 8 weeks practice and reliability training	\$40 per hour plus travel. Contact author: rholland-coviello@air.org	1 day	None
Classroom Observation of Early Mathematics Environment and Teaching (COEMET)	Unspecified, but 2 days recommended to study and practice the instrument	Information Not Available	No less than half a day	Contact Dr. Sarama: jsarama@buffalo.edu, 716-645-1155
Child Observation Form and Scale (COFAS)	Approximately 1 week (½ day training plus 3 – 4 days practice and reliability training)	Information Not Available	Approximately 20 minutes per caregiver	None
Classroom Practices Inventory (CPI)	Several weeks of preliminary observations and practice	No Formal Training	At least 2.5 hours	Information Not Available
The Emergent Academic Snapshot (EAS)	Training by authors	Information Not Available	1 – 2 hours for one child; 3 – 6.5 hours for snapshot of up to 4 children	Information Not Available
Early Childhood Classroom Observation Measure (ECCOM)	2 day training	Contact Dr. Deborah Stipek at stipek@stanford.edu	3 hours	Contact Dr. Deborah Stipek: stipek@stanford.edu
Early Childhood Environment Rating Scale Extension (ECERS-E)	½ day training plus 2 guided observations	Varies: \$620-\$820 per person for basic training; \$2500 per person for training on research standards	At least ½ day	\$27 plus shipping for scales
Early Childhood Environment Rating Scale – Revised (ECERS-R)	3 – 5 day training including 2 practice observations	\$59 for video training package, \$4 for training workbook	At least 3 – 3.5 hours	\$19.95 for manual

	Training		Administration	
	<i>Training Needed</i>	<i>Cost of Training</i>	<i>Time Needed for Administration</i>	<i>Cost of Measure</i>
Early Language and Literacy Classroom Observation (ELLCO)	Minimum of 9 hours	Information Not Available	1 – 1.5 hours	\$50 for User's Guide and Toolkit
Early Language & Literacy Classroom Observation Pre-K Tool (ELLCO – Pre-K)	Contact Brookes Publishing: brookespublishing.com, 800-638-3775	Contact Brookes Publishing: brookespublishing.com, 800-638-3775	Approximately 3 hours	\$50 for User's Guide and Toolkit
Early Language & Literacy Classroom Observation: Addendum for English Language Learners (ELLCO: Addendum for ELL)	Contact Dina Castro at castro@mail.fpg.unc.edu	Contact Dina Castro at castro@mail.fpg.unc.edu	1 – 1.5 hours	Free from author (Dina Castro) after training completed
Early Literacy Observation Tool (ELOT)	3 step training: read manual, formal training, practice sessions	Varies: Usually \$2000 for up to 40 people plus trainer travel expenses plus \$20 per person for materials	At least 80 minutes	\$350 per school for processing results submitted online; \$450 per school for processing results submitted on scantron sheets
Emlen Scales	No training required	N/A, no training required	Approximately 10 minutes	Cost of mailing/distributing questionnaires, data entry and analysis
Environment and Policy Assessment and Observation (EPAO)	1 day	Contact Dr. Diane Ward: dsward@email.unc.edu	1 day	Contact Dr. Diane Ward: dsward@email.unc.edu
Family Child Care Environment Rating Scale – Revised Edition (FCCRS-R)	Formal group training or self-training with video training package	\$59 for video training package	3.5 hours	\$16.95 for manual with score sheet

	Training		Administration	
	<i>Training Needed</i>	<i>Cost of Training</i>	<i>Time Needed for Administration</i>	<i>Cost of Measure</i>
Individualized Classroom Assessment Scoring System (inCLASS)	2 days	Contact University of Virginia trainers	Approximately 4 hours/1 morning	Information Not Available
Infant and Toddler Environment Rating Scale – Revised (ITERS-R)	Group or self-training with video training package	\$59 for the video observations training package	At least 3 hours	\$16.95
Language Interaction Snapshot (LISn)	2 days plus 4 hours of individual study	Depends on number of participants (contact Mathematica) (609) 799-3535 www.mathematica-mpr.com	15 minutes per child observed	None
Observation Measures of Language and Literacy Instruction (OMLIT)	Classroom training (8 hours for each of the 4 central measures, less than ½ day for OMLIT-CLOC and Classroom Description) plus practice observation	Varies on number of measures to be trained: \$1000 a day plus travel expenses for a central measure training	At least 3 hours	Contact Abt Associates for PDFs of each measure: www.abtassociates.com
Observational Record of the Caregiving Environment (ORCE)	2 day training plus practice	Information Not available	1.5 – 3 hours	None
Program Administration Scale (PAS)	4 trainings available depending on goal. Trainings range from 1 - 2 hours to 4 days.	Contact Jill Bella at the McCormick Center for Early Childhood Leadership: (800) 443-5522, ext. 5059 jill.bella@nl.edu	4 – 6 hours	\$19.95
Preschool Classroom Implementation Rating Scale (PCI)	On-going during use: 2 days – 2 weeks	Information Not Available	At least 1 day per visit	Information Not Available
Program for Infant/Toddler Care Program Assessment Rating Scale (PITC PARS)	At least 3 day training plus practice	Contact WestEd, Center for Child and Family Studies Evaluation Team: www.WestEd.org	At least 3 hours for observation component	Varies – contact authors for options

	Training		Administration	
	<i>Training Needed</i>	<i>Cost of Training</i>	<i>Time Needed for Administration</i>	<i>Cost of Measure</i>
Preschool Mental Health Climate Scale (PMHCS)	Information Not Available	Information Not Available	Information Not Available	Information Not Available
Preschool Program Quality Assessment Instrument (PQA)	2 day training	Information Not Available	At least 1 day	\$25.95
Preschool Rating Instrument for Science and Math (PRISM)	At least 5 days training	Varies; researchers should be trained by the developers. Contact Dr. Ellen Frede, efrede@nieer.org	Full session for half-day program, 4 hours from start of day for full-day program	Measure not publicly available.
Quality of Early Childhood Care Settings: Caregiver Rating Scale (QUEST)	2.5 day training	\$2500 plus expenses for up to 10 participants	Information Not Available	Information Not Available
Ramey's Observation of the Learning Environment (ROLE)	1 – 2 weeks training and practice	Information Not Available	2 – 3 hours	None
Ready School Assessment (RSA)	Ready School Teams attend 2 day workshops; customized training and technical assistance available	Contact HighScope: infor@highscope.org 734-485-2000	Varies	\$199.95 for 5 copies of instrument, 1 Administration Manual, 5 Team Handbooks, 5 Questionnaires, 2 year license for online profiler
School Age Care Environment Rating Scale (SACERS)	3 or 5 day training	\$1225 per person for 5 day training, \$825 per person for 3 day training	2.5 hours	Information Not Available
Supports for Early Literacy Assessment (SELA)	Several hours for discussion and inter-rater reliability training	Information Not Available	3 – 3.5 hours	Contact author: Dr. Sheila Smith, Sheila.Smith@nyu.edu, 212-998-5014
Supports for Social-Emotional Growth Assessment (SSEGA)	Several hours for discussion, practice, and inter-rater reliability training	Information Not Available	3 – 3.5 hours	Contact author: Dr. Sheila Smith, Sheila.Smith@nyu.edu, 212-998-5014

	Training		Administration	
	<i>Training Needed</i>	<i>Cost of Training</i>	<i>Time Needed for Administration</i>	<i>Cost of Measure</i>
Teacher Behavior Rating Scale (TBRS)	At least 2 days (2 weeks for those without early education experience)	\$1890 for training at UTHSC-Houston (plus travel expenses for local training)	2 – 3 hours	\$35 for manual
Teacher Instructional Engagement Scale	No training has been developed	No Formal Training	At least 1 hour	Free with request of author
Teacher Knowledge Assessment (TKA)	No training required	No Formal Training	Approximately 45 minutes	Free with request of authors
Teaching Pyramid Observation Tool for Preschool Classrooms (TPOT)	Information Not Available	Information Not Available	At least 2 hours	Information Not Available

The format and content of this table is specific to *Quality in Early Childhood Care and Education Settings: A Compendium of Measures, Second Edition* (Halle, Vick Whittaker, & Anderson, 2010) and was compiled by the authors.

Table 6: Technical Glossary

	What is it?	Why is it useful?
Score Distributions	The dispersion or spread of scores from multiple assessments for a specific item or scale.	In order for items and scales (sets of items) to be useful, they should be able to distinguish differences between programs. If almost every program scores low on a particular scale, it may be that the items make it "too difficult" to obtain a high score and, as a result, don't distinguish between programs on this dimension very well.
Inter-rater Reliability	How much assessments by different trained raters agree when observing the same program at the same time	It is important to use instruments that yield reliable information regardless of the whims or personalities of individual observers. If findings depend largely on who is rating the program (rater A is more likely to give favorable scores than rater B), it is hard to get a sense of the program's actual strengths and weaknesses.
Internal Consistency	The cohesiveness of items forming an instrument's scales	Scales are sets of items within an instrument that jointly measure a particular concept. If, however, the items within a given scale are actually conceptually unrelated to each other, then the overall score for that scale may not be meaningful.
Criterion Validity	When a variable or set of variables predicts some measure of performance or criterion.	It is useful to know whether an instrument, particularly quality measures, predict performance measures (e.g., children's math or reading scores). Some researchers suggest that concurrent and predictive validity are types of criterion validity.
Construct Validity	When an instrument measures the construct that it is intended to measure.	It is helpful to know exactly which concepts an instrument is measuring. There is not a statistical test which measures construct validity, so the best method is to compare your measure to other measures of the same construct, or parts of the same construct.
Concurrent Validity	When an instrument compares favorably with a similar measure (preferably one with demonstrated validity strengths).	It is important to use an instrument that generates accurate information about what you are trying to measure. If two instruments are presumed to measure similar concepts, one would expect findings from each instrument to be similar.
Discriminant Validity	When an instrument correlates or varies inversely with an accepted measure of the opposite construct.	Like concurrent validity, discriminant validity allows you to determine whether your instrument measures the construct that it is purported to measure. If two instruments are presumed to measure opposite constructs, one would expect these measures to vary inversely.
Predictive Validity	When an instrument successfully predicts related outcomes.	The best way to know whether a quality assessment instrument generates accurate information about what you are trying to measure is to see whether programs that score high on quality actually produce better outcomes for the children participating in the program.
Content Validity	When an instrument can be generalized to the entire content of what is being measured.	It is important that a measure capture an entire construct rather than only parts, or pieces of the construct. Although there may be disagreement about what constitutes a construct, it is important that measures try to assess all aspects.

The format for this table is from *Measuring Youth Program Quality: A Guide to Assessment Tools* (Yohalem & Wilson-Ahlstrom with Fischer & Shinn, 2009) and was used with permission from Nicole Yohalem. In addition, we reproduced, with permission, the exact wording of definitions for the following terms from *Measuring Youth Program Quality: A Guide to Assessment Tools*: Score Distributions, Inter-rater Reliability, Concurrent Validity, and Predictive Validity. Definitions for the following terms were produced specifically for *Quality in Early Childhood Care and Education Settings: A Compendium of Measures* (Halle & Vick, 2007) by the authors: Criterion Validity, Construct Validity, Discriminant Validity, and Content Validity.

Assessment Profile for Early Childhood Programs (APECP) Assessment Profile for Family Child Care Homes (APFCCH)

I. Background Information

Author/Source

Summative Measure:

- *Assessment Profile for Early Childhood Programs: Research Edition II*
- *Assessment Profile for Homes with Young Children: Research Version*
- *Assessment Profile for Early Childhood Programs: Research Edition Technical Manual*

Publisher:

Martha Abbott-Shim
294 Woodview Drive
Decatur, GA 30030
Email: martha.abbottshim@gmail.com

Formative Measure:

- *Assessment Profile for Early Childhood Programs*
- *Assessment Profile for Family Child Care Homes*

Publisher:

Quality Assist, Inc.
17 Executive Park Drive, Suite 150
Atlanta, GA 30329
Phone: 404-325-2225
Website: www.qassist.com

Purpose of Measure

Summative Measure: The Assessment Profile for Early Childhood Programs (APECP): Research Edition II is a global measure of quality used by researchers to evaluate the learning environment and teaching practices in classrooms for young children. The Assessment Profile for Homes with Young Children: Research Version was developed using items on the Assessment Profile for Early Childhood Programs: Research Version. The Family Child Care Homes version has only been used in the NICHD Early Childhood Research Project and the authors have never established any psychometric properties.

Formative Measure: The Assessment Profile for Early Childhood Programs: Pre-school, Toddler, Infant, School-Age, and Administration instruments are formative evaluation measures used for program improvement purposes. These measures are more comprehensive than the summative, research tool and provide user-friendly procedures for self-evaluation of early childhood settings. As formative measures, they are supported by software that provides extensive analyses and detailed program improvement recommendations. The Assessment Profile for Early Childhood Programs tool evaluates center-based, classroom and administrative practices while the Assessment Profile for

Family Child Care Homes (APFCCH) is a companion tool for formative evaluation purposes in the family child care setting. From 1988 to 1999 the National

Association of Family Child Care Homes (NAFCC) had exclusive rights for use of this instrument in their accreditation process.

Population Measure Developed With

The Assessment Profile for Early Childhood Programs: Research Edition I (1992) was originally standardized using 401 pre-school classrooms in child care, Head Start, and kindergarten settings. In 1998 the authors revised the instrument using a national standardization sample of 2,820 classrooms: 190 Head Start classrooms in two southern states, and 933 kindergarten, 935 first grade and 762 second grade classrooms across 31 states and the Navajo Nation. Subsequent analyses across the original 87 items were conducted to confirm the factor structures, to estimate reliability, and to recalibrate the IRT properties. Following the analyses, each scale was reduced to 12 items and the Assessment Profile: Research Edition II was published in 1998. The psychometric properties are reported in the Technical Manual. The Research Edition II of the Assessment Profile has been used in a number of research studies; for more information contact Dr. Martha Abbott-Shim (Martha.abbotshim@gmail.com).

Age Range/Setting Intended For

Summative Measure: The APECP: Research Edition II is an appropriate measure of classrooms for children 3-7 years of age.

Formative Measure: The APECP evaluates administrative practices as well as infant (birth-12 months), toddler (12-26 months), pre-school (3-5 years) and school-age (5-10 years) classrooms. The APFCCH evaluates mixed-age groups and therefore includes infant through school-age children as well as the business and professional practices of the family child care provider.

Ways in which Measure Addresses Diversity

Developmental diversity among children is addressed in both summative and formative measures through a series of criteria that focus on child assessment and individualizing instruction. In addition, the formative measure includes criteria under Curriculum Methods that address cultural diversity as the teacher incorporates a variety of language, customs and traditions (including food, music, art, stories, etc.) into the classroom environment and activities.

Key Constructs & Scoring of Measure

Summative Measure: The APECP: Research Edition II is an observation checklist with dichotomous items and includes five scales with 12 items each to assess: Learning Environment, Scheduling, Curriculum Methods, Interacting, and Individualizing. These five scales have met the unidimensionality criteria for Information Response Theory (IRT) creation of scales and have shown a strong fit to a three-parameter IRT model (Abbott-Shim, Neel, & Sibley, 2001).

Formative Measure: The APECP evaluates the Safety (109 items), Learning Environment (73 items), Scheduling (34 items), Curriculum Methods (49 items), Interacting (61 items), and Individualizing (25 items) practices within classrooms. The number of items for each dimension vary depending upon the age group observed; the maximum number of items are noted in parentheses. Administrative practices are evaluated in terms of Physical Facilities (68 items), Food Service (45 items), Program Management (63 items), Personnel (38 items), and Program Development (31 items).

The APFCCH evaluates Safety (51 items), Health and Nutrition (60 items), Learning Environment (41 items), Interacting (51 items), Outdoor Environment (25 items), and Professionalism (50 items).

The following Table provided by the developer presents a description of the constructs for each of these dimensions.

ASSESSMENT PROFILE FOR EARLY CHILDHOOD PROGRAMS

ASSESSMENT PROFILE FOR FAMILY CHILD CARE HOMES

CLASSROOM / HOME SETTINGS

SAFETY & HEALTH

Safety and Health focuses on the maintenance of a healthy and safe classroom environment with specific attention to the handling of emergency situations, basic health care and medication needs. Diapering procedures will be assessed in the infant and toddler classrooms.

LEARNING ENVIRONMENT

The Learning Environment focuses on the availability and accessibility of a variety of learning materials to children in the classroom. Variety is assessed across various conceptual areas (such as science, math, language, fine motor, etc.) and within a conceptual area. In addition, the arrangement of the classroom and outdoor space are assessed to determine if it encourages child independence and focused learning.

SCHEDULING

Scheduling assesses the Teachers' intentional planning as well as the implementation of classroom activities. Scheduling is assessed in terms of balance in variety of learning contexts (individual, small group, and large group) and learning opportunities (child-directed and teacher directed, quiet and active, indoor and outdoor experiences).

CURRICULUM METHODS

Curriculum Methods focuses on the variety of teaching techniques and strategies used to facilitate children's learning. Curriculum Methods also examines the opportunities for emergent learning, for children to guide their own learning, and for cooperative learning experiences.

INTERACTING

The interactions between the Teacher(s) and children are observed to assess the child's experience with positive physical and verbal interactions, Teachers' responsiveness, and approaches to behavior management.

INDIVIDUALIZING

Individualizing assesses the Teachers' implementation and use of systematic and comprehensive child assessment in planning and organizing learning experiences that match the skill level of each child. It also assesses the Teachers' system for identifying and providing for children with special needs and the Teachers' system for routine communication with parents.

SAFETY

The provision of a safe environment is essential to the quality of the child's care and focuses on the general condition of the home with specific consideration to the play areas, bathroom, diapering area, kitchen, and sleeping areas. Also reviewed is the Provider's ability and preparedness to handle emergency situations, critical to ensuring the safety of children.

HEALTH AND NUTRITION

The provision of basic health care and the encouragement of personal hygiene are two important aspects of high quality child care. Provision of basic health care includes the Provider's awareness of childhood illnesses, policies for handling illness, and a readiness to respond to health problems. Family child care Providers are responsible for providing nutritionally balanced foods, providing for individual differences, and encouraging sound nutritional habits while creating an atmosphere that encourages the development of social skills at mealtimes.

LEARNING ENVIRONMENT

The arrangement of the family child care learning environment impacts the quality of the child's experience. The home should be arranged to promote child independence, foster the child's sense of belonging, and provide a variety of activities and opportunities that meet the varying developmental needs of the children in care.

INTERACTING

The way in which the Provider interacts with the child influences the child's overall development and experience. Effective Providers have the ability to initiate warm and affectionate relationships with the children; facilitate learning; respond to children's needs; and effectively manage children's behavior.

OUTDOOR ENVIRONMENT

The outdoor environment is viewed as an extension of the child's overall learning environment. Providers should maintain a safe and healthy outdoor environment and provide opportunities for a variety of play and learning activities.

ASSESSMENT PROFILE FOR EARLY CHILDHOOD PROGRAMS

ASSESSMENT PROFILE FOR FAMILY CHILD CARE HOMES

ADMINISTRATION	PHYSICAL FACILITY The Physical Facility dimension focuses on the safe and healthy conditions of the indoor and outdoor physical facilities with specific considerations to the toileting and hand washing facilities and vehicle safety.
	FOOD SERVICE The Food Service dimension assesses the administrator's responsibility for providing menus, which reflect a comprehensive food service and nutritionally balanced diet, that will meet the individual needs of the children. Also assessed are the food handling procedures and the food preparation area.
	PROGRAM MANAGEMENT Program management is a review of the comprehensive, descriptive documentation of policies and procedures for staff and parents. Specific consideration will be paid to medication policies and program record keeping.
	PERSONNEL The Personnel dimension focuses on the administrator's responsiveness and support for the staff. Assessed is the administrator's ability to facilitate staff cohesiveness and positive working relationships in a program staffed with qualified individuals.
	PROGRAM DEVELOPMENT Program development focuses on the professionalism of the program's administrator, the system of evaluation for the staff and program and the professional development opportunities available to the staff.

PROFESSIONALISM

As a family child care professional, the Provider enters into a partnership with parents for the care and education of their children. It is the responsibility of the Provider to support this partnership through effective communication and clarification of policies and procedures. In addition, a high quality Provider demonstrates commitment to ongoing personal and professional growth.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Data collection requires observation, review of records, and interview with teachers, administrator(s), and/or family child care provider(s).

Training Required: Training is required to establish inter-rater reliability. Training involves a review of the criteria and data collection methods and on-site practice observation, record review, and interviews. Training generally involves 2 – 3 days.

Setting

Center-based pre-school classrooms	Assessment Profile for Childhood Programs: Research Edition II
Center-based	Assessment Profile for Early Childhood Programs
Family Child Care Homes.....	Assessment Profile for Family Child Care Homes

Time Needed and Cost

The time required to complete the *Assessment Profile* varies with each setting.

Center-Based Programs2 – 3 classrooms can be evaluated in approximately one day and involves morning observations in classrooms, afternoon review of records and teacher interviews.
For the formative evaluation tool only, the administrative component requires approximately 4 – 6 hours.

Family Child Care Homesapproximately 4 – 6 hours

Summative Measure: APECP: Research Edition II: \$18 (3 classrooms), Technical Manual: \$25.

Formative Measure: APECP and APFCCH pricing are based on the scope and specification of the evaluation plan regarding training, data collection, technological support (PDA), data analysis and reporting.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

For both the summative and formative versions of the Assessment Profile, inter-rater reliabilities between a trainer and observers is consistently reported with a mean of 93

to 95% agreement with a range of 83 to 99% agreement (Abbott-Shim, Lambert, & McCarty, 2000).

Internal Consistency

The reliability coefficients for the five scales (Learning Environment, Scheduling, Curriculum, Interacting, and Individualizing) range from .79 to .98 for the Kuder-Richardson 20 and from .81 to .98 for the Spearman-Brown corrected split-half. The IRT based reliabilities for the five scales range from .83 to .91 (Abbott-Shim, Neel & Sibley, 1992).

Validity Information

Criterion Validity

Criterion related validity was established by examining the relationship of the Assessment Profile: Research Edition I to the Early Childhood Environment Rating Scale (ECERS) (Harms & Clifford, 1980). In these criterion related validity studies, Wilkes (1989) found a significant correlation ($r = .64, p < .001$), and Abbott-Shim (1991) found a significant correlation ($r = .74, p = .001$).

Construct Validity

A second-order factor analysis was used to determine whether the five scales of the Assessment Profile: Research Edition II form a single latent construct of classroom quality. The path coefficients for each of the Scales are reasonably similar between Year 1 and Year 2 (Learning Environment .41 and .37; Scheduling .31 and .34; Curriculum .69 and .59; Interacting .59 and .52; Individualizing .45 and .59). The goodness of fit indices for the two years are as follows: root mean square residual .034 (Yr. 1) and .038 (Yr. 2); goodness of fit .99 (Yr. 1) and .99 (Yr. 2); adjusted goodness of fit .95 (Yr. 1) and .96 (Yr. 2); and normed fit index .96 (Yr. 1) and .93 (Yr. 2). These results indicated that observed measurements using these factor scores stem from a single underlying construct of classroom quality (Abbott-Shim, Lambert, & McCarty, 2000).

Content Validity

Content validity was documented through a review of the instrument by a wide range of early childhood professionals and a cross-reference of the items with the initial NAEYC Accreditation Criteria (National Association for the Education of Young Children, 1998). The cross-reference showed extensive consistency between the two measures with 100% match of the criteria. This cross-reference has been periodically updated as the accreditation criteria have been modified (Abbott-Shim, Neel, Sibley, 2001).

References and Additional Resources

Abbott-Shim, M. (1991). *Quality care: A global assessment*. Unpublished manuscript. Georgia State University.

Abbott-Shim, M., Lambert, R., and McCarty, F. (2000). Structural model of Head Start classroom quality. *Early Childhood Research Quarterly*, 15(1), 115-134.

- Abbott-Shim, M., Neel, J., and Sibley, A. (2001). *Assessment Profile for Early Childhood Programs- Research Edition II: Technical Manual*. Atlanta, GA: Quality Counts, Inc.
- Abbott-Shim, M., and Sibley, A. (1992). *Assessment Profile for Early Childhood Programs: Research Edition I*. Atlanta, GA: Quality Assist, Inc.
- Abbott-Shim, M., and Sibley, A. (1998). *Assessment Profile for Early Childhood Programs: Research Edition II*. Atlanta, GA: Quality Counts, Inc.
- Abbott-Shim, M., and Sibley, A. (1987). *Assessment Profile for Early Childhood Programs: Preschool, Toddler, Infant, School Age and Administration*. Atlanta, GA: Quality Assist, Inc.
- Lambert, R., Abbott-Shim, M., & Sibley, A. (2006). Evaluation the Quality of Early Childhood Education Settings. In B. Spodek & O. N. Saracho (Eds.) *Handbook of Research on the Education of Young Children, Second Edition*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.
- Wilkes, D. (1989). Administration, classroom program, sponsorship: Are these indices of quality care in day care centers. (Doctoral Dissertation, Georgia State University, 1989). *Dissertation Abstracts International 50*, AAI8922912.

Assessment of Practices in Early Elementary Classrooms (APEEC)

I. Background Information

Author/Source

Source: Hemmeter, M. L., Maxwell, K. L., Ault M. J., & Schuster J. W. (2001). *Assessment of Practices in Early Elementary Classrooms (APEEC)*. Teachers College Press: New York, NY.

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

Purpose of Measure

As described by the authors:

"The National Association for the Education of Young Children (NAEYC) position statement on developmentally appropriate practices (DAP) applies to children birth-through eight years of age (Bredekamp & Copple, 1997). However, most existing measures of DAP. . .focus on children from birth through kindergarten. . .The APEEC was developed to provide a useful tool for both practitioners and researchers who want to understand elementary school practices (K-3) in general education classrooms serving children with and without disabilities. The APEEC does not measure specific curriculum content or in-depth teacher-child interactions" (Hemmeter, Maxwell, Ault, & Schuster, 2001, p. 1).

Population Measure Developed With

- 60 professionals were contacted to participated in the review process: 30 faculty members and 30 practitioners.
- 46 (77%) of professionals returned completed interviews: 25 (83%) practitioners, and 21 (70%) faculty members. As a result of feedback, the measure was reduced from 40 items to 22 items.
- Interrater agreement and validity data on the revised 22 item measure was collected in 38 K-3 classrooms in 1997. As a result of low interrater agreement, the measure was further reduced from 22 items to 16 items.
- The final 16-item measure was field-tested in 69 classrooms in North Carolina and Kentucky in the spring of 1998.

Age Range/Setting Intended For

As described by the authors:

"The APEEC was designed to measure practices in K-3 general education classrooms that include children with disabilities for at least part of the day. However, it may also be used in classrooms with only typically developing children. Because the APEEC contains items measuring practices for children with disabilities, alternative scoring instructions are given for these items if no children with disabilities are

served in the classroom. The APEEC . . . does not measure aspects of the broader school environment, such as the playground or special subject classes" (Hemmeter et al., 2001, p. 3).

Ways in which Measure Addresses Diversity

Diversity is defined broadly in this measure to include gender, disability, family configurations and languages/cultures. The measure was designed for use in classrooms serving children with disabilities for at least part of the day, so issues related to diverse learners are incorporated into multiple items. Two particular diversity items are:

- Item 12 (Observation and interview): Rates the participation of children with disabilities in classroom activities, assessing the extent to which children with disabilities participate in many of the same classroom activities as children without disabilities and the extent to which IEP objectives are addressed through regular classroom activities.
- Item 14 (Observation and interview): Rates the degree to which materials and information on diversity are present in the classroom, and the extent to which diversity is discussed or integrated in the classroom and daily activities.

Key Constructs & Scoring of Measure

The APEEC consists of 16 items covering three broad domains of classroom practices: physical environment, curriculum, and instruction. All items are rated on a seven point likert-type scale. "A score of '1' indicates the classroom is *inadequate* in terms of developmentally appropriate practices, a score of '3' indicates *minimal* developmentally appropriate practices, a score of '5' indicates the classroom is *good* in terms of developmental appropriateness, and a score of '7' indicates *excellent* developmentally appropriate practices. Intermediate scores of '2', '4', and '6' can also be obtained." (Hemmeter et al., 2001, p. 4)

Descriptors are provided at points 1, 3, 5 and 7. Ratings are made using information collected both through classroom observation and teacher interview, with more weight placed on classroom observation.

- *Physical Environment* (4 items)
 - Room Arrangement
 - Display of Child Products
 - Classroom Accessibility
 - Health and Classroom Safety
- *Instructional Context* (6 items)
 - Use of Materials
 - Use of Computers
 - Monitoring Child Progress
 - Teacher-Child Language

- Instructional Methods
- Integration and Breadth of Subjects
- *Social Context* (6 items)
 - Children's Role in Decision Making
 - Participation of Children with Disabilities in Classroom Activities
 - Social Skills
 - Diversity
 - Appropriate Transitions
 - Family Involvement

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The APEEC should be administered by individuals knowledgeable about developmentally appropriate practices, early elementary classrooms, and special education practices. Individuals are expected to familiarize themselves with the items and scoring procedures and to read over the administration instructions provided by the authors.

Training Required: Observers should be trained to criterion prior to using the instrument. This requires that at least two people observe in the same classroom at the same time and rate the classroom independently. Inter-rater agreement should be at least 80% within 1 point.

Setting

The APEEC measures practices in K-3 general education classrooms. Because of its focus on the classroom setting, the APEEC does not measure aspects of the broader school environment, such as the playground or special subject classes (e.g. physical education, music, art).

Time Needed and Cost

Time: As this measure is based largely on classroom observations, the authors recommend observing as much of a full day's in-class activities as possible. This is followed by a 20-30 minute interview with the teacher.

Cost: \$13.95 (paperback)

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Inter-rater agreement data were available for 59 classrooms. At the item level, the average percentage of exact agreement was 58% (range: 31% - 81%), and the average percentage of agreement within 1 point was 81% (range: 50% - 100%). The median

weighted kappa was 0.59. Weighted kappas were 0.50 or higher for 12 items, and 2 items had a weighted kappa below 0.47.

Internal Consistency

The intraclass correlation (ICC) between the two observer's ratings was 0.86.

Validity Information

Construct Validity

Construct Validity was established by comparing the APEEC to several measures of developmentally appropriate practices. Correlations with each scale are presented below.

- The Assessment Profile for Early Childhood Programs (Abbott-Shim & Sibley, 1988), $r = 0.67$
- The Teacher Beliefs and Practices Scale (Buchanan, Burts, Bidner, White, & Charlesworth, 1998; Charlesworth, Hart, Burts, Thomasson, Mosley, & Fleege, 1993)
 - Developmentally appropriate practices, $r = 0.55$
 - Developmentally inappropriate practices, $r = -0.28$
- The Caregiver Interaction Scale (Arnett, 1989), $r = 0.61$

Comments

For a description of the APEEC in K-3rd classrooms, see:

Maxwell, K. L., McWilliam, R. A., Hemmeter, M. L., Ault, M. J., & Schuster, J. W. (2001). Predictors of developmentally appropriate classroom practices in kindergarten through third grade. *Early Childhood Research Quarterly*, 16, 431-452.

References and Additional Resources

Abott-Shim, M., & Sibley, A. (1988). *Assessment Profile for Early Childhood Programs*. Atlanta, GA: Quality Assist.

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10, 541-552.

Bredenkamp, S., Copple, C. (Eds.). (1997). *Developmentally appropriate practice in early childhood programs* (rev. ed.). Washington, DC: National Association for the Education of Young Children.

Buchanan, T. K., Burts, D. C., Bidner, J., White, F., & Charlesworth, R. (1998). Predictors of the developmentally appropriateness of the beliefs and practices of first, second, and third grade teachers. *Early Childhood Research Quarterly*, 13, 459-483.

Charlesworth, R., Hart, C. H., Burts, D. C., Thomasson, R. H., Mosley, J., & Fleege, P.O. (1993). Measuring the developmental appropriateness of kindergarten teachers. *Early Childhood Research Quarterly*, 8, 255-276.

Hemmeter, M. L., Maxwell, K. L., Ault M. J., & Schuster J. W. (2001). *Assessment of Practices in Early Elementary Classrooms (APEEC)*. Teachers College Press: New York, NY.

Maxwell, K. L., McWilliam, R. A., Hemmeter, M. L., Ault, M. J., & Schuster, J. W. (2001). Predictors of developmentally appropriate classroom practices in kindergarten through third grade. *Early Childhood Research Quarterly*, 16, 431-452.

Business Administration Scale for Family Child Care (BAS)



I. Background Information

Author/Source

Source: Talan, T. N. & Bloom, P. J. (2009). *Business Administration Scale for Family Child Care*. New York, NY: Teachers College Press.

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

Purpose of Measure

As described by authors:

"The Business Administration Scale for Family Child Care was designed to serve as a reliable and easy-to-administer tool for measuring and improving the overall quality of business practices in family child care settings. The content of the BAS reflects the wisdom in the field about the components of high-quality family child care. High-quality programs are run by providers who are intentional in their work with children and families, committed to ongoing professional development, engaged in ethical practice, and savvy about accessing community resources to enhance the effectiveness of their programs. High-quality programs have business practices and policies in place that promote financial stability, reduce the risk associated with doing business in a home environment, and comply with local and state legal requirements" (Talan & Bloom, 2009, p. 1).

The BAS is applicable for multiple uses, including program self-improvement, technical assistance and monitoring, training, research and evaluation, and public awareness. The BAS was designed to complement the Family Child Care Environment Rating Scale-Revised (FCCERS-R; Harms, Cryer, & Clifford, 2007). Both instruments measure quality on a 7-point scale and generate a program profile to guide program improvement efforts. When used together, these instruments provide a comprehensive picture of the quality of the family child care learning environment and the business practices that undergird the program.

Population Measure Developed With

An initial reliability and validity study of the Business Administration Scale for Family Child Care was conducted in early 2007 with 64 family child care providers in Illinois. Data generated from this initial sample were used to make revisions in the wording of different indicators, delete redundant items, and streamline the data-collection protocol. The refinements from this initial study resulted in the current version of the BAS.

The sample for the reliability and validity study of the current version of the BAS was drawn from 83 family child care providers in Florida, Tennessee, California, and Illinois. These states were selected as they varied in the stringency of state licensing regulations that govern family child care and provided a diverse national sample of providers. Thirty percent of providers were located in California, 32% in Florida, 30% in Tennessee, and 8% in Illinois. Providers were located in urban, suburban, and rural geographic regions of their state. Average BAS scores did not vary as a function of a family child care home's geographic location. At the item level, programs in California scored somewhat higher on "Income and Benefits" compared to programs in Florida, and programs in California and Tennessee scored higher on "Provider-Parent Communication" than did programs in Florida and Illinois.

In each of these states, a local quality improvement technical assistance agency was contacted to assist with data collection. Individuals with expertise in early childhood education were trained to administer the BAS. Technical assistance agencies were asked to recruit family child care programs that ranged in size and in quality.

The BAS requires that providers document many business practices considered personal in nature. Consequently, the sample drew from providers who had previously established relationships with their local technical assistance agencies. These providers were assumed to be more willing to provide documentation and to participate in quality improvement activities than providers who had no prior relationships with their local technical assistance agencies. While this sample does not reflect the overall population of providers in the United States, it does represent the providers who are most likely to use the BAS.

Of the family child care programs participating in the reliability and validity study, 12% were considered "small" and served 1 to 5 children, 45% were considered "medium" sized and served 6 to 10 children, and 43% were considered "large" and served 11 to 16 children. The average licensed capacity of the programs in the sample was 11.6 children. Providers actually enrolled an average of 8.3 full-time children and 1.5 part-time children. Approximately 97% of providers enrolled infants and toddlers, 94% enrolled pre-schoolers ages 3 to 4 years, and 61% enrolled school-age children ages 5 to 12 years. Of the 606 children served, 34% were infants and toddlers, 42% were pre-schoolers, and 24% were school-aged children. Average total BAS scores and BAS item scores did not vary as a function of the size of the family child care program.

Age Range/Setting Intended For

The BAS is intended for use in family child care programs that serve children of various ages. Average BAS scores do not vary as a function of the age groups that providers' serve.

Ways in which Measure Addresses Diversity

The BAS was designed to be applicable for use in family child care programs of varying size and in different geographic regions of the United States. The instrument acknowledges the diversity of family child care providers in the item Marketing and Public Relations, as it measures providers' participation in a variety of community-based organizations, such as a church, synagogue, mosque, Rotary International, and Chamber of Commerce. The instrument also addresses the diversity of families in the item Provider-Parent Communication. Specifically this item requires that, "The provider speaks the parent's primary language or utilizes resources in a family's primary language to communicate," and more generally the item is constructed so that providers have flexibility to implement communication strategies that are meaningful to the population of families they serve.

Key Constructs & Scoring of Measure

The BAS contains 37 indicator strands clustered in 10 items, which are rated on a 7-point scale from inadequate to excellent. The items include:

- *Qualifications and Professional Development*
- *Income and Benefits*
- *Work Environment*
- *Fiscal Management*
- *Recordkeeping*
- *Risk Management*
- *Provider-Parent Communication*
- *Community Resources*
- *Marketing and Public Relations*
- *Provider as Employer*

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The instrument can be administered by family child care providers as a self-assessment for quality improvement or by technical assistance specialists, educators, researchers, and program evaluators for accountability, monitoring, quality improvement, and research purposes.

Training Required: For self-assessment, no training is required. Measurement reliability and validity information is not available under this type of test administration. For accountability, monitoring, and research purposes, test administrators are required to participate in a two-day training on the instrument and are required to code a videotaped interview and review program documentation to achieve reliability.

Setting

The BAS was designed to assess the quality of business practices in family child care programs.

Time Needed and Cost

Time: Approximately one hour to interview the provider and an additional hour for a review of documents.

Cost: \$19.95

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Inter-rater reliability was determined during a two-day training on the use of the instrument with 21 assessors. Using videotaped interviews and a review of sample documentation for the entire data collection process, assessors were rated on how often they matched the BAS anchor's scores within 1 point on each item. Individual assessor inter-rater reliability scores ranged from 90% to 100% agreement on the overall BAS score. Overall average inter-rater reliability for the 21 assessors was 94%. Individual item reliability scores ranged from 67% to 100% with the median item reliability score calculated at 100% agreement.

Internal Consistency

The BAS items are organized under one common factor. A Cronbach's Alpha was conducted to determine how well the set of items measured the unidimensional construct. Coefficient alpha for the total 10-item scale ($n = 65$) was calculated at .77, and for the 9-item scale ($n = 83$) at .73, indicating that the BAS has acceptable internal consistency among items and that the items reliably measure the construct (only 9 items are used when the provider does not employ any assistants).

Validity Information

Criterion Validity

Criterion validity for the BAS was determined by a correlational analysis with one subscale of the Family Child Care Environment Rating Scale-Revised (FCCERS-R) that measures organizational effectiveness—the Parents and Provider subscale. Correlations between the overall BAS score and the FCCERS-R Parents and Providers subscale score yielded a significant and moderate, positive ($r = .49$) relationship suggesting that the BAS measures related, but not redundant, characteristics of organizational quality as those measured by the FCCERS-R.

Concurrent Validity

To establish the BAS's concurrent validity, an analysis of variance (ANOVA) was conducted to determine if higher BAS scores were related to higher global family child care quality using the FCCERS-R. Providers were grouped into those who scored at or below 3.50 on the FCCERS-R (the mid-point of the scale) and those who scored higher than 3.50. F tests ($f = 6.103, p = .019$) revealed that lower global

quality providers scored significantly lower ($M = 3.08$, $SD = .89$) on the BAS, than providers who had higher global quality ($M = 3.87$, $SD = .94$).

Content Validity

Content validity for the BAS for Family Child Care was established by a panel of seven early childhood experts who evaluated each item and indicator to ensure that key business management practices of a family child care program were included. Content reviewers were asked to respond to the following questions and provide feedback:

- Do the items cover the most important areas of business management in family child care settings?
- Do the indicators under each item adequately represent each item?
- Do the indicators appropriately show increasing levels of quality on a continuum?
- Does the wording of the item headings adequately reflect their content?

Multiple refinements were made to the wording and layout of the BAS as a result of the feedback provided by the reviewers. Additional revisions were made from feedback received by assessors who collected data in the initial reliability and validity study and through data analysis. As a result, the wording and order of indicators was changed and redundant items removed to assure that the BAS was applicable to a full range of family child care programs.

Comments

The sample from the reliability and validity study produced a mean score of 3.78 (on a scale of 1 to 7) with a standard deviation of 1.03. Scores ranged from 1.88 to 6.40, indicating an acceptable distribution of item scores and overall BAS scores across the quality continuum. On average, providers scored highest on the item Work Environment, and lowest on the item Fiscal Management.

References and Additional Resources

- Harms, T., Cryer, D., & Clifford, R. M. (2007). *Family Child Care Environment Rating Scale – Revised Edition (FCCERS-R)*. New York: Teachers College Press.
- Talan, T.N. & Bloom, P.J. (2009). *Business Administration Scale for Family Child Care*. New York, NY: Teachers College Press.

Classroom Assessment of Supports for Emergent Bilingual Acquisition (CASEBA)

I. Background Information

Author/Source

Source: Freedson, M., Figueras-Daniel, A., & Frede, E. (2009). Classroom Assessment of Supports for Emergent Bilingual Acquisition. New Brunswick, NJ: National Institute for Early Education Research.

Publisher: The CASEBA is still under development and is not currently publicly available. Please contact Dr. Ellen Frede for further information (efrede@nieer.org).

Purpose of Measure

The CASEBA is "designed to assess the degree to which pre-school teachers and classrooms are providing support for the social, cognitive, and linguistic development of English Language Learners (ELLs), with a focus on language and literacy" (Freedson, Figueras-Daniel, & Frede, 2009, p.1).

Population Measure Developed With

The development and validity testing of the CASEBA was completed on a statewide sample of publicly supported pre-school classrooms in New Jersey. These include public school, child care and Head Start programs. The classrooms were randomly selected and all included at least some children who speak Spanish at home. The instrument is also currently being used in a statewide study of pre-school effects in New Mexico and in a study comparing the effects of English speaking versus Spanish speaking teachers on DLL development.

Age Range/Setting Intended For

The CASEBA was designed for settings with English Language Learner pre-school students. Subscales of the instrument are also suitable to assess supports for language and literacy for all pre-school children.

Ways in which Measure Addresses Diversity

The CASEBA is designed for settings with English Language Learner pre-school students and assesses the teachers' cultural responsiveness.

Key Constructs & Scoring of Measure

The instrument includes 26 rating scale items which cluster around six broad aspects of the early childhood curriculum:

- *Collection of child background information* (2 items)
- *Supports for home language development* (11 items)
- *Supports for English acquisition* (8 items)



- *Social-emotional supports and classroom management* (3 items)
- *Curriculum content* (1 item)
- *Assessment* (1 item)

"Each of the 26 items measures one component of a high-quality classroom environment and instruction based on research about effective language and emergent literacy supports for 3- to 5- year old children who speak a language other than English at the home, and who are in the process of acquiring English as a second language. Each item is rated on a 7-point Likert scale, where 7 indicates that a specific form of support and accompanying practices are present in close to an ideal form, while 1 represents the total absence of any such practices" (Freedson, Figueras-Daniel & Frede, 2009, p.1).

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The CASEBA is designed to be used by researchers.

Training Required: Researchers should be trained to reliability by the developers.

Setting

The CASEBA is carried out in pre-school classrooms where students are English Language Learners.

Time Needed and Cost

Time: Training requires at least four days. In a half-day classroom administration is for the complete session. In a full-day classroom, the observation begins before children arrive and ends at nap time.

Cost: Varies.

III. Functioning of Measure

Reliability Information

Not yet available.

Validity Information

Not yet available.

Comments

The CASEBA is currently undergoing research to determine the psychometric properties of the instrument, including concurrent and predictive validity.



References and Additional Resources

Freedson, M., Figueras-Daniel, A., & Frede, E. (2009). *Classroom Assessment of Supports for Emergent Bilingual Acquisition: Overview and instructions for use*. New Brunswick, NJ: National Institute for Early Education Research.

Freedson, M., Figueras-Daniel, A., & Frede, E. (2009). *Classroom Assessment of Supports for Emergent Bilingual Acquisition*. New Brunswick, NJ: National Institute for Early Education Research.

The Child Care Assessment Tool for Relatives (CCAT- R)

I. Background Information

Author/Source

Source: Porter, T., Rice, R., & Rivera, E. (2006). *Assessing quality in family, friend and neighbor care: The child care assessment tool for relatives*. New York, NY: Institute for a Child Care Continuum.

Publisher: This measure is currently unpublished.

Purpose of Measure

As described by the authors:

This "observation instrument is specifically designed for measuring quality in child care provided by relatives" (Porter et al., 2006, p. i).

"The CCAT-R consists of five components: the Action/Communication Snapshot; the Summary Behavior Checklist; the Health and Safety Checklist; the Materials Checklist and the Caregiver Interview" (Porter et al., 2006, p. 9).

Population Measure Developed With

"The field test for the CCAT-R began in early 2004. . . A total of 92 observations were completed with caregivers in low-income communities in California, Arizona, Chicago and New York City. Fifty-two percent of the caregivers were Latino, 26% European American, and 21% African American. The remainder self-identified as other ethnic groups. The vast majority of caregivers were women (96%), but there were four men. More than half (55%) were grandparents of the children in care, and slightly more than a third (36%) were aunts or uncles. Approximately 9% were related to the children in some other way, such as cousins. The majority of the caregivers (61%) were married or living with a significant other. Among those who were single heads of households, half were never married; the remainder were separated, divorced or widowed.

Although no questions about income were asked, it is likely that most of the caregivers had low incomes, because they were recruited in low-income neighborhoods. Slightly more than half (58%) were paid for providing child care. Of the 53 caregivers who were paid, 31 received payment from the parent, 19 received payment from the government, and 3 received payment from both the government and parents. Approximately 70% of the caregivers who responded to the interview question about payment indicated that they could afford to provide care without it. Half of them said that parents gave them gifts or performed some service in exchange for the care.

There was a wide range of educational levels among participants. Of those caregivers who reported this information, 40% had high school degrees or equivalent, and another 45% had some college, a two-year degree or a four-year degree. Approximately 15% of the caregivers had not completed high school.

There was also a wide range of child care experience. Approximately 13% had been caring for children for a year or less, 44% of the caregivers had five or fewer years of experience providing child care for other people's children, and nearly 20% had been taking care of children for 20 years or more.

Caregiver's training in early childhood education varied, too. Slightly more than half (53%) of the caregivers had some sort of specialized training such as Child Development Associate classes, teacher training, nurse's training, child care workshops, parent education workshops, or some other type of training (e.g., training for foster care). Nearly a quarter of the sample had taken classes in child development or early education at a college or university.

On average, caregivers provided care for two children; the range of children in care varied from one to seven. Approximately 40% of the arrangements consisted of one child; another 21% provided care for two children. Slightly more than a third of the children (38) were under three. Approximately 16% of the caregivers indicated that they were caring for children with special needs such as attention deficit hyperactivity disorder, learning delays, or asthma" (Porter et al., 2006, p.15-16).

Age Range/Setting Intended For

The CCAT-R was designed for settings in which a relative cares for children under age six.

Ways in which Measure Addresses Diversity

The sample in the field test included African American, Latino, and European American parents. Since the field test, the CCAT-R has been used successfully with Asian Americans as well (Porter & Vuong, 2008).

Key Constructs & Scoring of Measure

There are five constructs related to the caregiver's support for different developmental domains:

- *Support for physical development, including health and safety*
- *Support for cognitive development*
- *Support for language development*
- *Support for social/emotional development*
- *Relationship with parents*

The first four constructs are captured in the Action/Communication Snapshot and the Summary Behavior Checklist. The Health and Safety Checklist as well as the Materials Checklist identify practices and materials that are related to these constructs

as well. The Caregiver Interview is the only component that includes items related to relationships with parents, although it also includes items that are potentially related to support for other domains.

Caregivers are rated on four factors: caregiver nurturing; caregiver engagement in activity with child; caregiver/child bidirectional communication; and caregiver unidirectional use of language. Each is related to different constructs. The caregiver nurturing factor measures the caregiver's support for social/emotional development, while the caregiver engagement factor measures interactions that promote physical and cognitive development. Two factors relate to language: 1) Caregiver/child bidirectional communication, which reflects interactions around language between the caregiver and the child; and 2) Caregiver unidirectional use of language, which measures the caregiver's talk to the child. For each factor, summary scores are calculated by obtaining the average of the Snapshot and Behavior Checklist subtotals on individual items. There are two sets of scores on each factor: one for children under three years of age, the other for children three to five years of age.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The test can be administered by individuals with early childhood, parent education, or family support training. No specific educational level beyond high school is required. Test administrators must be able to speak and read English, because the CCAT-R is not available in other languages at this time.

Training Required: 2.5 day training is offered through The Institute for a Child Care Continuum at Bank Street College of Education or the organization's site.

Setting

Relative's home/child's home (setting where the relative provides the care for the child). Since its development, the CCAT-R has also successfully been used in group settings for family, friend and neighbor caregivers (Porter & Vuong, 2008).

Time Needed and Cost

Time: Observation: 2 – 2 ½ hours
Interview: 30 minutes

Cost: The cost for the training depends on the number of trainers needed. Cost per participant is \$1075. The training includes the instrument, the manual and a training DVD.

A scoring software program is available for \$250. Timing files, which are necessary for the timing intervals that are used in the Action/Communication Snapshot and the Summary Behavior Checklist, the two components that use time-sampling, are available for MP-3 players.

Information about the CCAT-R can be downloaded from the following website:
<http://www.bankstreet.edu/gems/ICCC/CCATRfinal5.8.06.pdf>

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

"Observers were trained to a criterion of .80 exact agreement on individual items in the CCAT-R Action/Communication Snapshot and the Summary Behavior Checklist in a minimum of 4 of the 6 observation cycles. Inter-rater reliability was obtained through comparison of observers' coding with the master-coded videotaped practice observation and two live observations with a reliable observer before observers used the CCAT-R in the field" (Porter et al., 2006, p. 16).

Validity Information

Criterion Validity

"It is possible that the CCAT-R has criterion validity as well because the items are grounded in child development theory and research and, as a result, may be predictive of positive child outcomes" (Porter et al., 2006, p. 16-17).

Construct Validity

"Initial confirmatory factor analysis indicated that there were too few cases of several items – toileting, for example – for statistically useful variation, and these items were eliminated. Subsequent analyses indicated that there were not enough unique items in behavior management to support it as a construct. We eliminated it from scoring, but retained the Behavior Checklist items in the coding for future research.

Additional confirmatory factor analysis using a maximum likelihood fit test with both promax and oblimin rotations produced five factors that seemed feasible. To check for consistency, we ran the generalized least squares fit function with promax rotation. Although the two solutions differed in several ways, there was satisfactory substantive correspondence in the first four factors to justify their use. . .

Some items, particularly those related to nurturing such as kissing, holding and patting, loaded on more than one factor, specifically the language factors. This may reflect the caregiver's interactions with infants and toddlers, because caregivers may hold babies as they talk to them. In addition, some of the caregiver talk items load on both language factors. The primary difference between these factors is that the child responds to the caregiver's talk in bidirectional communication, in the unidirectional use of language. In other words, the former measures caregiver talk with the child, while the latter measures caregiver talk to the child" (Porter et al., 2006, p. 16-17).

Concurrent Validity

The field test did not include a formal concurrent validity component, but it did compare four items from the Family Day Care Rating Scale (FDCRS) with the four

CCAT-R factors. With the exception of the FDRCS item, "Tone," scores were similar—that is, the median FDCRS score corresponded to the CCAT-R rating (Porter et al., 2006, p. 21).

Predictive Validity

The predictive validity of the CCAT-R has not been tested, but it is currently being used in a three-year longitudinal study of a cohort of 3-year-olds in a family interaction program in Hawai'i. The study is examining the relationship between quality in relative child care and child outcomes. Time 1 and Time 2 results will be available fall, 2009.

Content Validity

"The content validity is based on participation of child care researchers throughout the CCAT-R's development. A group of researchers reviewed the constructs that informed the individual items in the CCAT-R, and several reviewed the full CCAT-R before the pilot test. In addition, [the developers] discussed the constructs and the CCAT-R items with practitioners at national conferences to identify whether the measure reflected caregiver behaviors with which they had experience" (Porter et al., 2006, p.16).

Comments

Since its development, the CCAT-R has been used in several assessments of child care quality. They include the Early Head Start Enhanced Home Visiting Pilot Evaluation (Pausell, Mekos, Del Grasso, Rowand, & Banghart, 2006); and two small evaluations of state funded-CCDF initiatives for family, friend and neighbor caregivers in Alabama and New Mexico (Porter, 2005). It was also used in a pre/post evaluation of a family interaction program in Hawai'i with sample of 58 caregivers, many of whom are Native Hawaiian (Porter & Vuong, 2008). Additional evaluations in Los Angeles, CA, San Jose, CA, Chicago, IL, and Tempe, AZ are expected to use the CCAT-R as well.

References and Additional Resources

- Pausell, D., Mekos, D., Del Grasso, P., Rowand, C., & Banghart, P. (2006). *Strategies for supporting quality in kith and kin child care: Findings from the Early Head Start Enhanced Home Visiting Pilot Program evaluation Final report*. Princeton, NJ: Mathematica Policy Research, Inc.
- Porter, T. (2005). *Evaluating quality in family, friend and neighbor child care: Results from two case studies*. Presentation at the National Association of Child Care Resource and Referral Agencies Annual Conference, Washington, DC.
- Porter, T., Rice, R., & Rivera, E. (2006). *Assessing quality in family, friend and neighbor care: The child care assessment tool for relatives*. New York, NY: Institute for a Child Care Continuum.

Porter, T. & Vuong, L. (2008). *Tutu and me: Assessing the effects of a family interaction program on parents and grandparents*. New York: Bank Street College of Education.

The Child Care HOME Inventories (CC-HOME)

I. Background Information

Author/Source

Source: Bradley, R. H., Caldwell, B. M., & Corwyn, R. F. (2003). The Child Care HOME Inventories: Assessing the quality of family child care homes. *Early Childhood Research Quarterly, 18*, 294-309.

Adapted by: Maxwell, K. L. & Kraus, S. (2002). Child Care Home Inventory-Phone. FPG Child Development Institute, UNC-CH.

Publisher: Available online
www.fpg.unc.edu

Purpose of Measure

As described by the authors:

"The Child Care HOME Inventory (CC-HOME) was designed to measure the quality and quantity of stimulation and support available to a child in non-parental child care arrangements taking place in home-like settings other than the child's own home" (Bradley et al., 2003, p. 297).

The CC-HOME encompasses two measures: the Infant-Toddler Child Care HOME (IT-CC-HOME) and the Early Childhood-Child Care HOME (EC-CC-HOME).

Many of the existing measures that assess quality of care in family child care homes (e.g., FDCRS, ITERS, PROFILE, CIS, AIS, and ORCE) "have acceptable to good psychometric qualities, but most require quite extensive periods of observation and some require substantial training to use. Some, like the FDCRS and the PROFILE, focus primarily on the physical, instructional, and organizational features of the child care arrangements, whereas others (e.g., ORCE, the Arnett, AIS) concentrate primarily on interactions between caregiver and child" (Bradley et al., 2003, p. 295). "There is a need for valid, reasonably comprehensive measures of the quality of care individual children receive in family child care settings that can be given in a relatively brief visit to the informal care environment. That is the niche CC-HOMEs are designed to fill" (Bradley et al., 2003, p. 296).

The instrument was developed as part of the NICHD Study of Early Child Care (NICHD Early Child Care Research Network, 1996). The CC-HOME is suitable for research and evaluation purposes. The CC-HOME is also relevant for public policy purposes, as this tool may help licensing workers and others responsible for maintaining quality in child care to obtain useful information about family child care homes.

Population Measure Developed With

75% of the caregivers in home-like settings (other than care provided by relatives) agreed to participate in observations of the child care environment in the NICHD Study of Early Child Care. Those agreeing to participate were more likely to be of higher education and less likely to be African American. The IT-CC-HOME was used with in-home caregivers of 377 24-month-old children. The EC-CC-HOME was used with 274 caregivers of 3-year-olds. The children observed were primarily European American (88%), in nuclear families (approximately 70%) whose fathers lived with them (approximately 80%). About 10% of the child sample received public assistance.

Age Range/Setting Intended For

The Child Care HOME Inventory (CC-HOME) encompasses two measures: the Infant-Toddler Child Care HOME (IT-CC-HOME) designed for use when children are less than 3 years old, and the Early Childhood-Child Care Home (EC-CC-HOME) designed for use when children are 3-6 years old.

Settings appropriate for the CC-HOME include care by relatives and neighbors (outside of the child's home) as well as care in licensed and unlicensed family child care homes.

Ways in which Measure Addresses Diversity

According to the author, the CC-HOME was designed so that it could be used with a wide variety of families. The parent instrument on which the CC-HOME was modeled has been used in studies involving every major ethnic group in the U.S. and scores on the HOME generally correlate with measures of family and child functioning. The correlations do tend to be a little stronger for European American families than for other ethnic groups but meaningful correlations tend to be obtained within most every group.

Key Constructs & Scoring of Measure

The IT-CC-HOME is composed of 43 binary-choice items organized into six subscales:

- *Caregiver Responsivity* (11 items)
- *Acceptance* (7 items)
- *Organization* (6 items)
- *Learning Materials* (9 items)
- *Caregiver Involvement* (6 items)
- *Variety of Stimulation* (4 items)

The EC-CC-HOME is composed of 58 items organized into eight subscales:

- *Learning Materials* (11 items)
- *Language Stimulation* (7 items)
- *Physical Environment* (7 items)
- *Caregiver Responsivity* (8 items)

- *Academic Stimulation* (5 items)
- *Modeling of Social Maturity* (7 items)
- *Variety in Experience* (9 items)
- *Acceptance of Child* (4 items)

"There is considerable overlap in the content of the Infant/Toddler and Early Childhood version but the content of each version is targeted to the developmental needs of children within the age ranges specified" (Bradley et al., 2003, p. 299).

Comments

The IT-CC-HOME and EC-CC-HOME inventories are very similar to versions of the HOME (Home Observation of the Measurement of the Environment) Inventory used to assess the family environment (Caldwell & Bradley, 2003). There is over 90% overlap between CC-HOME and the original HOME for each age group. Minor modifications were made to the HOME inventories to make them appropriate for evaluating family child care environments. Specifically, the IT-CC-HOME contains 43 items, rather than the 45 items in the Infant/Toddler HOME. The EC-CC-HOME contains 58 items rather than the 55 in the Early Childhood HOME. "This close modeling results in nearly equivalent measures of environmental quality for family child care and the home environment for studies where measuring both environments is deemed desirable" (Bradley et al., 2003, p. 297).

The CC-HOME does not provide as intensive a level of coverage of caregiver-child interactions as do other measures (e.g., ORCE, Arnett, and the AIS), nor does it capture aspects of either the social or the physical environment in as much detail as the PROFILE and the FDCRS. The CC-HOMEs do not attempt to directly assess formal curricula. Instead of providing a deep or intensive coverage of any aspect of care, it provides broad coverage of the structural, organizational, and educational features of the caregiving environment.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Trained observers conduct the observations. For the NICHD Study of Early Child Care, about ½ day of training was required, followed by practice with the instrument, and achieving 90% reliability with criterion coding of videotaped child care settings.

Training Required: The authors note that it is "not generally necessary to have such intensive training in order to achieve reliability on the CC-HOME" (Bradley et al., 2003, p. 300).

Setting

Observations are made in family child care homes.

Time Needed and Cost

Time: The CC-HOME observation takes about one hour in the home-like child care setting.

Cost: Manual \$30.00
50 Forms \$25.00

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Paired observers went to each child care setting at each time point in the NICHD Study of Early Child Care. Each member of the pair scored each item on the CC-HOME independently, and their scores were compared using Pearson correlations and a repeated measures ANOVA procedure developed by Winer (1971).

At the 24-month data collection, 53 pairs of scores were examined. Pearson correlations ($r = .94$) and the Winer correlation ($r = .97$) were both very high.

At the 54-month data collection, 23 pairs of observations revealed very high reliability using both Pearson correlations ($r = .98$) and the Winer correlation ($r = .99$). "Although these estimates of inter-observer agreement are quite high, they are consistent with a review of studies on the original HOME Inventories done by Bradley (1994) which showed that simple levels of agreement are typically in the 90-95% range" (Bradley et al., 2003, p. 301).

Internal Consistency

The 45 items on the IT-CC-HOME has a Cronbach's alpha of .81 (NICHD Early Child Care Research Network, 1996).

Validity Information

Criterion Validity

Criterion validity of the HOME is well-established. Studies have linked HOME scores to various aspects of child well-being, suggesting that it is related to cognitive, motor, and social outcomes as well as to growth and health (Bradley, 1994; Bradley, Corwyn, & Whiteside-Mansell, 1996). "Establishing the criterion validity of the CC-HOMEs per se is more difficult in that the quality of the home environment typically accounts for far more variance in child well-being than does the quality of child care environments (NICHD Early Child Care Research Network, 2003)" (Bradley et al., 2003, p. 305).

Construct Validity

The CC-HOME was designed after the version of the HOME created to measure family environments, which was based on a review of child development and family

theory, as well as empirical research on actions, objects, events, and conditions that are associated with aspects of child well-being.

Concurrent & Discriminant Validity

Scores on the CC-HOME show moderate relations with the sensitivity and stimulation composites from the Observation Record of the Caregiving Environment (ORCE) used in the NICHD Study of Early Child Care (.46 - .58). and the Abbot-Shimm Assessment Profile for Early (.57 - .69).

Convergent Validity

Scores on the CC-HOME were correlated with scores on the ORCE and PROFILE for the NICHD Study of Early Child Care sample.

Subscale scores from the IT-CC-HOME were significantly correlated with the caregiver sensitivity and cognitive stimulation composite variables from the ORCE (correlations ranged from $r = .15$ to $r = .61$). Caregiver Responsivity and Caregiver Involvement showed high correlations with the Sensitivity composite of the ORCE ($r = .61$ and $.59$, respectively); the Caregiver Involvement also showed moderate relations with the Stimulation composite ($r = .44$).

Subscale scores from the EC-CC-HOME also had significant correlations with the ORCE Sensitivity and Stimulation composites (correlations ranged from $r = .18$ to $r = .55$). Caregiver Responsivity was highly correlated with the ORCE Sensitivity composite ($r = .55$). Moderate correlations were found between Learning Materials, Academic Stimulation, and Variety of Experience subscale scores and the ORCE Stimulation composite ($r = .35$, $.35$, and $.37$, respectively).

The CC-HOME was also significantly correlated with the PROFILE at both time points (correlations ranged from $r = .21$ to $r = .69$). For the IT-CC-HOME, the two strongest correlations were for Learning Stimulation ($r = .51$) and Caregiver Involvement ($r = .62$). For the EC-CC-HOME, the strongest correlations were for Variety of Stimulation ($r = .53$), Learning Materials ($r = .47$) and Language Stimulation ($r = .45$).

Content Validity

Extensive and careful review of the literature has undergirded the development of both the HOME and the CC-HOME. "The content validity of the CC-HOME rests on the strength of those reviews together with consultations with professionals who deal with children and families" (Bradley et al., 2003, p. 305).

References and Additional Resources

Bradley, R. H. (1994). The HOME Inventory: Review and reflections. In H. W. Reese (Ed.), *Advances in child development and behavior*, Vol. 25, 241-288. Orlando, FL: Academic Press.

- Bradley, R. H., Caldwell, B. M., & Corwyn, R. F. (2003). The Child Care HOME Inventories: Assessing the quality of family child care homes. *Early Childhood Research Quarterly, 18*, 294-309.
- Bradley, R. H., Corwyn, R., & Whiteside-Mansell, L. (1996). Life at home: Same time, different places, and examination of the HOME Inventory in different cultures. *Early Education & Parenting, 5*, 251-269.
- Caldwell, B. M., & Bradley, R. H. (2003). *Home Observation for Measurement of the Environment: Administration manual*. Little Rock, AR: University of Arkansas.
- NICHD Early Child Care Research Network. (1996). Characteristics of infant child care: NICHD early child care research network. *Early Childhood Research Quarterly, 11*(3), 269-306.
- NICHD Early Child Care Research Network (2003). Families matter – Even for kids in child care (commentary). *Journal of Developmental and Behavioral Pediatrics, 24*, 58-62.
- Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York, NY: McGraw-Hill.

Child Caregiver Interaction Scale (CCIS)

I. Background Information

Author/Source

Source: Carl, B., Dissertation for Indiana University of Pennsylvania, 2007.

Publisher: This measure is currently unpublished.
For more information, contact author by email at B.E.Carl@iup.edu.

Purpose of Measure

As described by the authors:

A review of established child care interaction measures revealed that no one assessment device exists for measuring the interaction between a child care provider and children in multiple age groupings and settings, ranging from infancy through school age and including family child care homes. Most caregiver interaction scales remain limited to specific age groupings and therefore do not cover the age spectrum found in most child care facilities.

The CCIS is a valuable and much needed measurement tool to assess child caregiver interaction across age groupings and settings. This measure not only provides a scale that can be used for research purposes to compare child care quality, but also serves as a noteworthy tool for training and technical assistance. By helping child caregivers understand their strengths and areas most in need for improvement, the CCIS is a tool that can be used to improve quality child care.

Population Measure Developed With

Original items were developed by the author and based on the National Association for the Education of Young Children's (NAEYC) Developmentally Appropriate Practice (DAP). Each of the items and indicators were reviewed by ten early childhood professionals. Reviewers were asked to evaluate the clarity and conciseness of each item and indicator, and to identify awkward or confusing items.

Data collection for the pilot study was conducted in conjunction with the 2006 Keystone STARS Quality Study, administered through the Office of Child Development (OCD), Pennsylvania Department of Public Welfare. The data collectors gathering the pilot CCIS data simultaneously collected Environmental Rating Scale data for the Quality Study. Additional data were collected from child care providers who participated in training programs, including Mind in the Making (social/emotional training for the care provider) and Child Development Credential (targeted child education) Programs. The sampling frame for this study consisted of 223 child care providers throughout the Commonwealth. Data were collected on infant/toddler, pre-school, school age, and family child care settings.

Participants in the pilot study comprised a reasonably representative sample of the larger Keystone STARS Quality Study and also of the total child care facilities population in Pennsylvania. Additionally, the pilot study included other specific groups (infant/toddler care and school aged care) that were not included in the larger 2006 study.

Age Range/Setting Intended For

The measure is designed to assess interactions between caregivers and children in multiple age groupings, ranging from infancy through school age. The assessment takes place in both home and center based child care settings.

Ways in which Measure Addresses Diversity

Two items measure diversity in the classroom and are part of the *Social domain* (described below).

- "Engaging children with special needs": Rates the extent to which children with special needs are included in the group, the extent to which adaptations are made within the classroom to facilitate/enable children with disabilities to participate in classroom activities, how comfortable caregivers are interacting with and caring for children with special needs, and the extent to which caregivers are included as part of the IFSP/IEP.
- "Cultural competence": Rates the extent to which daily routines and classroom materials represent different races, cultures, ages, abilities and genders in non-stereotyping roles, and the extent to which staff intervene to counteract prejudice and promote understanding and acceptance of diversity.

Key Constructs & Scoring of Measure

The CCIS consists of 17-items, covering three domains: *Emotional, Cognitive/Physical, and Social*. Each item is assessed on a seven point scale ranging from 1 (inadequate) to 7 (excellent). Several indicators are available at anchor points 1, 3, 5 and 7.

- *Emotional domain* (6 items): tone of voice, acceptance/respect for children, greeting, enjoys and appreciates children, and expectations for children, health and safety.
- *Cognitive/Physical domain* (7 items): routines/time spent, physical attention, discipline, language development, learning opportunities, involvement with children's activities, symbolic and literacy materials.
- *Social domain* (4 items): promotion of prosocial behavior/ social emotional learning, engaging children with special needs, relationships with families, cultural competence.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The measure should be administered by a reliably trained, objective assessor.

Training Required: Training consists of a one day review of the scale, reviewing each item and indicator. A minimum of two follow up reliability observations are recommended to ensure accurate interpretation of the measure.

Setting

The CCIS can be administered with infant, toddler, pre-school, and school-aged after school caregivers, in both center- and home-based child care settings.

Time Needed and Cost

Time: It is recommended that a three hour block of time be used for the administration of this scale. Administration of this scale can be conducted simultaneously with the age/setting appropriate ERS.

Cost: Contact the author for training and use of the scale at B.E.Carl@iup.edu.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Inter-rater reliability was established between the author and two other observers prior to the start of data collection. Initial reliability proved a high percentage of agreement by each observer on each item within one point on the seven-point scale (95%). No items were off by more than one score point. Reliability was assessed independently with each of the two observers through the course of the study. A high level of inter-rater reliability was maintained by each of the observers, demonstrating an Inter Class Correlation (ICC) ranging from .88 to .93, with each item within one point on the seven-point scale.

Due to incomplete data collection the original set of 17 items was decreased to 15. Item #3, "Greeting," was omitted because of incomplete data. Item #15, "Engaging With Special Needs Children," was also omitted because of the low number of cases (n = 24) where a special needs child was enrolled in the program. Information presented is based upon the adjusted pilot sample of 181.

Internal Consistency

Cronbach's alpha for the CCIS measure, across all age groups and settings was .94. Cronbach's alpha for pre-school age caregivers was .95, for infant/toddler caregivers was .91, for home based caregivers was .93, and for school-aged caregivers was .95. Analyses of the theoretically derived subscales of emotional, cognitive and social each revealed a moderately high Cronbach's Alpha with relatively high corrected

item-total correlations. The Emotional subscale was comprised of 5 items ($\alpha = .87$), the Cognitive subscale consisted of 7 items ($\alpha = .88$), and the Social subscale consisted of only three items ($\alpha = .72$).

Validity Information

Concurrent Validity

Concurrent validity was explored by correlating the CCIS average and the age/setting appropriate overall Environmental Rating Scale (ERS) average, which were collected at the same time. The correlation between the CCIS and the overall ERS ratings average was significant ($.74, p < .001$).

Convergent & Discriminant Validity

Convergent validity was assessed by exploring the correlation between the CCIS average and the "Interaction" subscales of the age/setting appropriate ERS. This subscale was chosen for comparison because of its theoretical association with the CCIS in terms of caregiver interaction, versus a purer measure of the physical environment.

Discriminant validity was explored by assessing the correlation between the CCIS average and the "Space and Furnishings" subscale of the age/setting appropriate ERS. This was chosen for analysis because the "Space and Furnishings" subscale provides a stronger focus on the classroom environment versus that of the caregiver interaction.

The correlation between the CCIS and the "Interactions" subscale of the ERS scale was also significant, ($.75, p < .001$). Again, this indicates a moderate to strong positive linear relationship between the two assessment scales. However, while the correlation between the CCIS and the "Space and Furnishings" subscale of the ERS was significant, it was lower than the other two correlations ($.67, p < .001$).

Predictive Validity

For purposes of this analysis, the factors of education, STAR level (quality enhancement rating), years of experience in child care, and the adult/child relationship were explored using multiple regression. The multiple regression analysis revealed the linear combination of caregiver characteristics were significantly related to the CCIS score, $F(4, 146) = 4.85, p < .001$. Analysis revealed a statistically significant relationship between the education of the provider and the CCIS score irrespective of the other variables. It also indicates that after controlling for the other variables a statistically significant relationship exists between the STAR level of the child care facility and the CCIS score.

Content Validity

The CCIS is based upon the solid theoretical base of DAP and is structured to incorporate these principles. To ensure consistency between DAP and the CCIS, many item indicators of the CCIS include specific examples drawn from DAP. Further, the training materials for data collectors are directly drawn from DAP. This

attention to coordination between the DAP and data collection documents ensured the CCIS was built upon both research and theory, and ensured strong content validity.

Comments

Care should also be taken in the interpretation of the results of the CCIS. Feedback results on individual item responses are not advised. Because each of the items is combined with others to create a subscale for the cognitive, emotional and social domains, it is recommended that the lowest level of feedback provided to caregivers be on the domain level.

Practitioners also need to be clear on how each of the subscales combines to create an overall caregiver interaction score. Because of the interconnected nature of these domains, research from this study indicate that caregivers who scored high on one subscale also tended to score high on the others. Using the CCIS to help caregivers identify and target desired behavior can be a useful tool in increasing the quality of child caregiver interactions.

References and Additional Resources

Carl, B., Dissertation for Indiana University of Pennsylvania, 2007.

The Child-Caregiver Observation System (C-COS)

I. Background Information

Author/Source

Source: Boller, K., & Sprachman, S. and the Early Head Start Research Consortium (1998). *The Child-Caregiver Observation System Instructor's Manual*. Mathematica Policy Research, Inc: Princeton, NJ.

Publisher: Mathematica Policy Research, Inc.
P.O. Box 2393
Princeton, NJ 08543-2393
(609) 799-3535
Fax: (609) 799-0005
Website: <http://www.mathematica-mpr.com/>

Purpose of Measure

As described by the authors:

"C-COS is a child-focused observation system that captures the experiences of an individual child in a caregiving environment over a two-hour period using a time-sampling procedure" (Boller & Sprachman, 1998, p. 1). It was developed to allow for comparisons of the quality of care provided across setting type (centers, family child care homes). The language categories were adapted from the items in the Observation Record of the Caregiving Environment (ORCE; NICHD ECCRN, 1996) that were found to be most associated with children's language development.

Population Measure Developed With

The C-COS was developed for the Early Head Start National Research and Evaluation Project (EHSREP; U.S. Department of Health and Human Services, 2004). The EHSREP was implemented in 17 EHS programs in all regions of the country. Programs offered center-based, home-based, and mixed-approach services. The families and children who participated in the evaluation were diverse. Many of the families were single-parent, were ethnically diverse (including Hispanic, African American, and White), did not speak English as their primary language, had relatively low educational attainment, and were receiving public assistance of some kind (e.g., Medicaid, WIC, food stamps, AFDC or TANF, and SSI benefits). A total of 3,001 families participated in the evaluation, with 1,513 in the treatment group and 1,488 in the control group. The C-COS was developed for use and added to the child care quality assessments when children were 24 and 36 months old. While collecting C-COS data, field staff also rated programs using the appropriate version of the Environment Rating Scales (ITERS; Harms & Clifford, 1990, ECERS; Harms, Clifford, & Cryer, 1998, or FDCRS; Harms & Clifford, 1989) and the Arnett Caregiver Interaction Scale (Arnett, 1989). At 24 months, the C-COS was conducted in 387 center-based toddler classrooms and in 141 family child care homes. At 36

months, the C-COS was conducted in 488 center-based classrooms and in 99 family child care homes.

(http://www.acf.hhs.gov/programs/opre//other_resrch/eval_data/reports/common_constructs/com_ch3_pro_hseval.html)

Age Range/Setting Intended For

The C-COS is intended for use with one- to five-year-old children in all types of child care settings. The C-COS was also adapted for a study on children younger than one year old.

Ways in which Measure Addresses Diversity

The C-COS has been used in large studies of children from diverse racial, ethnic, linguistic, and economic backgrounds.

Key Constructs & Scoring of Measure

"The C-COS is conducted during a two-hour child care observation. Every 20 minutes, the observer begins a child-focused observation that lasts five minutes, during which the observer is prompted by an audiotape to observe the child for 20 seconds and record the codes on the coding sheet for 10 seconds" (Boller & Sprachman, 1998, p. 1).

- 'focus child' (FC) designates the child whose interactions and activities will be observed
- 'provider of care' (DP) describes the caregiver with primary responsibility for the focus child throughout the day

There are eight coding categories in the C-COS. The first five, labeled A through E on the C-COS form, are filled in during the 10-second record periods that occur throughout each five-minute child-focused observation.

A. Type of Caregiver Talk

- Responds to Focus Child (FC) talk
- Language or Communication Requested
- Action Requested
- Reading
- Other Talk/Singing

B. FC Talks to...

- Self or Unknown
- Other Children
- Direct Provider
- Other caregivers

C. FC Interaction With or Attending to...

- Other Child(ren) or Group
- Caregiver
- Material (Played with or explored)
- Television or Video
- None: Wandering/Unoccupied

D. FC was...

Smiling

- E. The Main Caregiver Interaction or Attempting to Interact with FC was...
Direct Provider of Care
Other Caregivers
All Caregivers Roughly Equal
No Interaction

The overall quality ratings, F through H, are completed at the end of the five-minute observation” (Boller & Sprachman, 1998, p.5). These are rated on a five point scale: (0) Ignoring/None; (1) All Negative; (2) Mostly Negative; (3) Mostly Positive/Neutral; (4) All Positive/Neutral.

- F. Caregiver Behavior towards FC
G. FC Behavior towards Caregiver
H. FC Behavior towards Other Children

Constructed variables that can be derived from the source data include the proportion of observed time that included "Any Caregiver Talk," "Caregiver Responding to Child," "Child Negative Behavior," "Focus Child Talk," Child Attending to Television/Video," "Child Wandering/Unoccupied," "Caregiver Initiating Talk with Child," "No Caregiver Interaction."

The C-COS can be used for sequential snapshots of one child or with a sample of children at each 5-minute period to get an overall measure of interaction and child activity.

II. Administration of Measure

Who Administers Measure/Training Required

Administration: Trained C-COS observers spend 2 to 3 hours in a setting observing and recording the target behaviors.

Training Required: Observers should be trained by a C-COS Instructor, practice the C-COS in the field in at least one child care center and one family child care home, and test their reliability by coding the C-COS test tape. The authors recommend one day of classroom training and two field observations to become familiar with the measure and establish reliability. The manual recommends that a new instructor on the C-COS spend approximately 2 days reviewing the training materials and practicing with videotapes. Prior to participating in a training session, trainees should read the manual, review the form, and complete the exercise on coding child care provider talk. The trainer should schedule approximately one hour for the lecture portion of the training and 3.5 hours for the coding practice portion of the training. Trainees conduct post-training activities on their own by viewing the training videotape and transcripts and becoming comfortable with the coding system. When ready, they conduct at least two practice observations in the field and then take the videotaped test. If a trainee does not pass, there are additional test tapes available.

Setting

The C-COS is designed for use in all types of child care settings.

Time Needed and Cost

Time: Approximately 2 hours per observation.

Cost: C-COS Instructor's Manual: Free. Copies of the audio prompts for the observing and recording periods and of the training and test videotapes: Approximately \$250.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

The Growing Up in Poverty study (2000) examined the impacts of welfare reform on children's early development in a sample of 800 children of welfare recipients, ages 30 – 42 months. The C-COS was used to examine child-caregiver interactions in center-based child care programs or pre-schools and licensed child care homes. In this study an inter-rater reliability on the C-COS of .90 was attained.

The Early Head Start Research and Evaluation Project (EHSREP; U.S. Department of Health and Human Services, 2004) is a large-scale randomized evaluation of Early Head Start (EHS). The Birth- Three Phase of the project (1996 – 2001) investigated the impact of EHS on children at three ages (14, 24, and 36 months). The C-COS was developed for this study and used to measure caregiver-child interactions. An intraclass correlation that surpassed the minimum of .80 was found for the C-COS in this study (U.S. Department of Health and Human Services, 2004).

The Who Leaves Who Stays Study (Phillips, Crowell, Whitebook, & Bellm, 2003) examined the literacy levels of early childhood educators in Alameda County, and how they are related to children's literacy environments. The sample included 98 teachers and their students in Head Start, public pre-schools, child care programs, and licensed family child care providers. The C-COS-Revised was used to measure caregivers' one-on-one interactions with children. As with the EHSREP, researchers found an intraclass correlation that surpassed the minimum of .80.

Internal Consistency

For constructs such as "Any Caregiver Talk," coefficient alpha ranged from .90 to .94 for the EHS REP sample (U.S. Department of Health and Human Services, 2004).

Validity Information

Concurrent Validity

In the EHSREP (U.S. Department of Health and Human Services, 2004), the C-COS construct Caregiver Talk was correlated with the ITERS-R (Harms, Cryer, and Clifford, 2003) and Arnett Caregiver Interaction Scale (CIS; Arnett, 1989) at 24 months (.24, $p < .01$ and .33, $p < .01$ respectively). At 36 months, Caregiver Talk

was not found to be correlated with the ECERS (Harms et al., 1998). In family child care, Caregiver Talk was correlated with the Arnett CIS (Arnett, 1989; $.23, p < .05$ at 24 months and $.34, p < .01$ at 36 months) but not with the ERS. This information can also be taken as evidence of discriminant validity.

In the Growing Up in Poverty Study (2000), C-COS as measured in center-based care was correlated with the Arnett CIS (Arnett, 1989; $r = .33$ to $.34$) and the ECERS (Harms et al., 1998; $r = .24$). In family child care the C-COS was associated with the Arnett CIS (Arnett, 1989; $r = .22$ -. $.29$). It was also correlated with caregiver education in family day care ($r = .29$). Urging focal child to talk correlated with mean ECERS scores (Harms et al., 1998; $r = .24, p < .002$). Frequency of wandering correlated with the total ECERS (Harms et al., 1998; $r = .54, p < .001$). Observed interactions between child and provider was associated with child care education ($r = .25, p < .03$).

In the Who Leaves Who Stays Study (Phillips et al., 2003), significant ($p < .01$) positive correlations were found between environmental quality (total ECERS, FDCRS, and ITERS scores) and items on the C-COS dealing with language interaction between providers/teachers and children and with children smiling and laughing. Significant ($p < .05$) negative correlations were found with environmental measures and C-COS items dealing with children being idle or upset. C-COS percent time child was wandering/unoccupied or watching television was correlated with percentage of staff with a BA or higher ($r = -.23, p < .05$). C-COS constructs were not significantly correlated with the percentage of staff with degrees in early childhood education nor with the percentage of highly trained staff (degree in ECE or more than 24 units). C-COS percent of time child attends to provider was correlated with the percentage of highly trained staff ($r = .23, p < .05$). The percentage of time the focus child smiles or laughs was negatively correlated with the percentage of highly trained staff ($r = -.25, p < .05$).

Comments

The C-COS was developed to provide a different perspective on quality than measures focused on the overall environment for children, with the hypothesis that the quality of care received by individual children may not be highly correlated with the overall quality of care. This makes assessing the validity of the C-COS challenging in that high correlations with overall quality would not be seen as a justification for the expense of conducting the observation. The ultimate test of the C-COS then is its predictive validity and those analyses have not been conducted. The EHSREP (U.S. Department of Health and Human Services, 2004) demonstrated that Early Head Start had a positive impact on C-COS scores in the areas of Caregiver Talk, Caregiver Responding to Child, and Caregiver Initiating Talk with Child. Thus, the C-COS is sensitive to differences in quality in the same way the ERS was found to vary across the EHS settings and the child care settings used by children in the control group.

References and Additional Resources

- Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10, 541-522.
- Boller, K., & Sprachman, S. and the Early Head Start Research Consortium (1998). *The Child-Caregiver Observation System Instructor's Manual*. Mathematica Policy Research, Inc: Princeton, NJ.
- Growing Up In Poverty Project (2000). *Remember the children: Mothers balance work and child care under welfare reform*. Growing Up in Poverty Project; Wave 1 Findings – California, Connecticut, Florida. Berkeley, CA: Graduate School of Education, PACE.
- Harms, T., & Clifford, R. M. (1989). *Family Day Care Rating Scale*. New York: Teachers College Press.
- Harms, T., & Clifford, R. M. (1990). *Infant/Toddler Environment Rating Scale*. New York: Teachers College Press.
- Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale*. Revised ed. New York: Teachers College Press.
- Harms, T., Clifford, R. M. & Cryer, D. (2005). *Early Childhood Environment Rating Scale – Revised Edition*. New York, NY: Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R. M. (2003). *Infant/Toddler Environment Rating Scale: Revised Edition*. New York, NY: Teachers College Press.
- NICHD Early Child Care Research Network (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly*, 11, 269–306.
- Phillips, D., Crowell, N., Whitebook, M., & Bellm, D. (2003). *Who leaves? Who stays? A longitudinal study of the early education and care workforce in Alameda County*. California Center for the Study of Child Care Employment. Berkeley: University of California.
- Porter, T., Rice, R., & Rivera, E. (2006). *Assessing quality in family, friend and neighbor care: The child care assessment tool for relatives*. New York, NY: Institute for a Child Care Continuum.
- U.S. Department of Health and Human Services. (2004, February). *The role of Early Head Start programs in addressing the child care needs of low-income families with infants and toddlers: Influences on child care use and quality*. Washington, DC: Department of Health and Human Services, Administration for Children and Families.

Child Development Program Evaluation Scale (CDPES)

I. Background Information

Author/Source

Source: Fiene, R. (1984). *Child Development Program Evaluation Scale and COFAS*. Washington, DC: Children's Services Monitoring Consortium.

Purpose of Measure

As described by the authors:

"The purpose in constructing the CDPE Scale was the perceived need in the child development program area to have a comprehensive scale that could be used by states or local agencies to determine compliance of child development programs with basic minimal requirements that ensure a child is in a safe and healthy environment" (Fiene, 1984, Introduction).

The scale also measures the quality of the child development program.

Population Measure Developed With

"The 37 item scale was selected from nearly 900 items. These 900 items were from different states' Compliance Instruments. . . It is a generic scale that incorporates results from Pennsylvania's Child Development Evaluation Instruments, West Virginia's and New York City's Child Development Compliance Instruments, California Child Development Quality Assessment Instrument, NAEYC and CWLA National Standards and the results of the National Day Care Study" (Fiene, 1984, Introduction).

Age Range/Setting Intended For

The CDPES may be used with infants, toddlers, pre-schoolers, and school-age children and is administered in the child care setting.

Ways in which Measure Addresses Diversity

One item assesses "Ethnic and Cultural Recognition" and evaluates the extent to which information is available to staff regarding traditional ethnic and cultural observances, learning opportunities are provided that acknowledge ethnic and cultural backgrounds of the children and community, activities are implemented to enhance a sense of cultural pride, each child shares his or her individual ethnic and cultural background, and staff provide multicultural experiences that broaden each child's knowledge of other cultures throughout the world.

Key Constructs & Scoring of Measure

The CDPES measures seven domains: administration, environmental safety, child development curriculum, health services, nutritional services, social services, and transportation.

Each domain is described in more detail below:

- *Administration* (6 items)
 - Staff qualifications
 - Adult child ratio/group size
 - Child development program
 - Employee performance evaluation
 - Personnel policies
 - Staff development
- *Environmental Safety* (4 items)
 - Whether the center is hazard free
 - Access to cleaning materials
 - Sufficient space
 - Equipment
- *Child Development Curriculum* (15 items)
 - Supervision of children
 - Observations about whether activities promote the development of skills, self-esteem, etc. (the Caregiver Observation Form and Scale (COFAS) is used to determine compliance with this item)
 - Goals and objectives
 - Identification of child's needs
 - Social emotional development
 - Physical development
 - Cognitive development
 - Language development
 - Art
 - Music
 - Dramatic play
 - Personal interaction
 - Self concept
 - Ethnic and cultural recognition
 - Special needs of the child
- *Health Services* (4 items)
 - Health appraisal
 - Emergency contact
 - Administration of medication
 - Child's health record
- *Nutritional services* (2 items)
 - Nutrition (in the licensing scale)
 - Nutrition (in the program quality scale)
- *Social Services* (5 items)

- Staff parent communication
- Family confidentiality
- Parent activities
- Parent involvement
- Parent education.
- *Transportation* (1 item)
 - Safety of the carrier.

While the CDPES may be used in its entirety to assess the seven domains above, it actually comprises two distinct scales: a center licensing scale and a program quality scale. Both are described in more detail below.

- *The Center Licensing Scale* assesses 13 items: health appraisal, caregiver observations, emergency contact, hazard free, cleaning materials, supervision of children, staff qualifications, group size and adult/child ratios, sufficient space, nutrition, administration of medication, safety carrier and equipment.
- *The Program Quality Scale* assesses the following items: child development program, employee performance evaluation, personnel policies, staff development, goals and objectives, identification of child's needs, social emotional development, physical development, cognitive development, language development, art, music and dramatic play, nutrition, personal interaction, self concept, ethnic and cultural recognition, special needs of the child, staff parent communication, child's health record, family confidentiality, parent activities, parent involvement, and parent education.

Items on the *Center Licensing Scale*, designed to rate compliance, are scored on a dichotomous scale, with a 0 indicating they are out of compliance and a 3 indicating that they are in compliance (there is no score of 1 or 2).

For the *Program Quality Scale*, observers rate on a score of 1 to 5, with 1 indicating the lowest quality, and a 5 indicating the highest quality. The program quality scale builds one level upon the other, such that in order to obtain a score of 3, the program must be doing everything at levels 1 and 2. For the majority of the questions, ratings can be determined by reviewing center documentation or interviewing staff members. Ratings of quality related to social-emotional development, physical development, cognitive development, language development, art, music, and dramatic play should be performed based on classroom observations.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The CDPES can be used by state licensing and monitoring staff, researchers, and directors of early care and education programs.

Training Required: Training on the CDPES requires 1-2 days of classroom training followed by on-site inter-rater reliability (usually 2-3 days). Individuals who are interested in using the scale should plan on 1 week of training and on-site implementation before using the scale for actual data collection.

Setting

The CDPES is administered in the child care setting. If there is more than one classroom in the center, one classroom is to be randomly selected and observations should be based on that classroom.

Time Needed and Cost

Time: Generally the CDPES can be completed in a day's time by one individual for programs that have fewer than 60 children. If the program is between 61-120 children, plan on 2 days to complete the scale and if 121 or greater plan on 3 days to complete the scale.

Cost: Free

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Inter-rater reliability kappa = .91

Internal Consistency

Cronbach's Alpha = .94 Total Scale

Validity Information

Construct Validity

Construct validity was assessed by comparing the CDPES with licensing and program quality assessment decisions and ratings ($r = .67; p < .01$).

Concurrent Validity

Concurrent validity was assessed by comparing the CDPES and the ECERS total scores ($r = .77; p < .005$).

Predictive Validity

"The licensing predictor items are statistically significant items that have been found to predict the overall compliance of child day care centers with state regulations in four states' regulations" (Fiene, 1984, Introduction).

Comments

The Caregiver Observation Form and Scale (COFAS) is used in conjunction with the CDPES to assess the behaviors of caregivers while interacting with children in a classroom setting.

The CDPES has been used in many states to assess the relationship between licensing and program quality. It was through these assessments that key licensing indicators were determined to distinguish between high quality programs. These results were published in several places, the most recent being the Office of the Assistant Secretary for Planning and Evaluation's "13 Indicators of Quality Childcare: Research Update 2002" (Fiene, 2002).

For additional information regarding the CDPES, please contact:

Richard Fiene, Ph.D., Associate Professor
Human Development and Family Studies
W-311 Olmsted Building
Penn State University - Harrisburg
777 West Harrisburg Pike
Middletown, Pennsylvania 17057
rjf8@psu.edu

References and Additional Resources

Fiene, R. (2002). *13 indicators of quality childcare: Research update*. Washington, DC: U.S. Department of Health and Human Services.

Fiene, R. (1984). *Child Development Program Evaluation Scale and COFAS*. Washington, DC: Children's Services Monitoring Consortium.

Child/Home Early Language & Literacy Observation (CHELLO)

I. Background Information

Author/Source

Source: Neuman, S., Dwyer, J., & Koh, S. (2007). *Child/Home Early Language & Literacy Observation Tool (CHELLO)*. Baltimore, MD: Brookes Publishing.

Neuman, S. B., Koh, S., & Dwyer, J. (2008). CHELLO: The Child/Home Environmental Language and Literacy Observation. *Early Childhood Research Quarterly*, 23, 159-172.

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

Purpose of Measure

As described by the authors:

"The CHELLO was created as an observational research tool to examine the physical and psychological environmental features of home-based child care associated with children's developing language and literacy skills. The CHELLO assesses the quality of early childhood language and literacy practice in family, friend and neighbor care settings. The CHELLO was designed to complement the Early Language and Literacy Classroom Observation (ELLCO) (Smith & Dickinson, 2002), which is an instrument used in center-based care settings. The CHELLO includes two research tools to assess instructional and affective supports in home-based care: 1) the Literacy Environment Checklist, used to assess the availability of resources and organization of space; and 2) the Group/Family Observation and Provider Interview, used to assess the instructional supports and affective environment for learning.

The CHELLO may be used for research and evaluation purposes including serving as a pre-assessment measure, as well as a tool for assessing intervention effects. The CHELLO can also be used as a professional development tool to improve the quality of the child care environment. The instrument also has the potential to be used for examining changes in home-based literacy interventions with parents. The CHELLO can be used in conjunction with the ELLCO to make comparisons between home-based and center-based care settings."

Population Measure Developed With

- Initial observations to develop the measure were conducted with 10 family/group centers recommended by a local resource and referral agency in Michigan.
- The final version of the measure was completed in spring 2005. This version was used in a study of 261 providers in four urban communities: Detroit, Flint, Grand Rapids, and Lansing Michigan (Project Great Start Professional Development Initiative).
 - All providers were female
 - Ethnically diverse: 10% Hispanic, 29% African-American, 59% White, 2% multi-racial
 - Average age was 39
 - Average child care experience: 10 years or less
 - Psychometric properties are based on the fall administration of the CHELLO on a sample of 119 home-based centers.

Age Range/Setting Intended For

The CHELLO was designed for use in mixed-age, home-based care settings.

Ways in which Measure Addresses Diversity

One item in the Adult Affect construct within the Support for Learning domain of the Group/Family Observation assesses the extent to which the provider "brings each child's home culture and language into the shared culture of the setting so that children feel accepted and gain a sense of belonging" (Neuman, Dwyer, & Koh, 2007 p.11).

Key Constructs & Scoring of Measure

The CHELLO is organized into three sections: a literacy environment checklist, a group/family observation form, and a provider interview. Each item within the Literacy Environment Checklist is rated as either Yes or No.

Each item within the Group/Family Observation is rated on a 5-point Likert-type scale: low (1), mid (2, 3) or high (4, 5). A rating of 1 is considered "Deficient," a rating of 3 is considered "Basic," and a rating of 5 is considered "Exemplary." Descriptors are provided at points 1, 3, and 5 on the scale.

Six provider interview questions supplement the information obtained from classroom observation elements. The interview items are particularly important to score features of the environment that may not be evident from a one-time observation (e.g., communication with parents).

The Literacy Environment Checklist contains 22 items addressing the following constructs:

- *Book Area* (4 items): Address the orderliness, comfort, and accessibility of an area set aside for reading books.

- *Book Use* (6 items): Address the number, types, and location of books in the child care environment.
- *Writing Materials* (6 items): Address materials available for writing (e.g., templates, tools, paper), whether there is a separate area set aside for writing, and whether the alphabet and children's writing are displayed in the setting.
- *Toys* (3 items): Address whether cognitively-stimulating toys, games/puzzles, and props to support dramatic play are available.
- *Technology* (3 items): Address the availability of computers, recorded books/stories, and other technology that supports children's language and literacy development (e.g., regularly watching the educational television program *Between the Lions*).

The Group/Family Observation contains 42 items reflecting 13 constructs organized into three domains:

- *Physical Environment for Learning*. This domain includes the following 3 constructs: Organization of the Environment (4 items), Materials in the Environment (4 items), and Daily Schedule (3 items).
- *Support for Learning*. This domain includes the following 3 constructs: Adult Affect (3 items), Adult-Child Language Interaction (4 items), and Adult Control Behaviors (3 items).
- *Adult Teaching Strategies*. This domain includes the following 7 constructs: Vocabulary Building (3 items), Responsive Strategies (3 items), Use of Print (3 items), Storybook/Storytelling Activities (4 items), Writing Activities (3 items), Monitoring Children's Progress (3 items), and Family Support and Interaction (3 items).

Summary scores for each construct are obtained. Subtotals are generated for the Literacy Environment Checklist, and the three domains from the group/family observation - Physical Environment for Learning, Support for Learning, and Adult Teaching Strategies. Finally, an overall CHELLO total score is obtained by summing the four subtotal scores.

Comments

Because the CHELLO was based on the ELLCO, there are some common items. Specifically, there are 19 items common to both measures. The Literacy Checklist contains 8 items that are the same as on the ELLCO, including book area, book availability, environmental print, and opportunities for children to write. The Observation and Interview has 11 items that are the same as on the ELLCO, including child choice, writing material and activities, reading materials and activities, organization of the environment for learning, parent communication, and ongoing progress monitoring. These subsets of items may be used for making comparisons across center-based and home-based settings, and may be particularly useful in contexts that involve multiple settings and multiple placements for children (Neuman, et al., 2007, p. 15).

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The instrument is administered by trained observers. It is recommended that CHELLO users have a strong background in language development and early literacy developmental practices. According to the authors (Neuman et al., 2007), Users might include:

- Researchers interested in assessing the quality of language and literacy experiences in early childhood in home-based settings
- Supervisors interested in improving the quality of home-based care
- Facilitators or coaches who want to target professional development efforts
- Program officers who wish to establish results-oriented accountability

Training Required: A manual, which provides observers with examples of how each item should be scored, should be reviewed prior to participating in the training session. The day-long training session includes a discussion (and examples) of each item. Observers view videotaped examples of exemplary, mediocre, and poor settings and use the instrument to rate the examples. Participants discuss explicitly how to score each section. Once the training has concluded, pairs of observers independently rate a home environment to establish reliability.

Setting

The instrument is meant to be used for observations of home-based child care settings (specifically, family, friend, and neighbor care).

Time Needed and Cost

Time: The entire instrument takes between 1 ½ to 2 hours to complete. The 22-item Literacy Environment Checklist was designed to take 10 minutes to complete.

Cost: Contact Susan Neuman: sbneuman@umich.edu

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Ten pairs of observers rated 20 home-based programs. There was a high rate of agreement between observers. The inter-rater reliability for the Literacy Checklist was 91%. The inter-rater reliability for the Observation section was 91% (within one point on a 1-5 scale).

In a follow up study including 128 home-based care settings, observers independently rated 30 home-based settings. Weighted kappas were calculated separately for the Literacy Environment Checklist and the Group/Family Observation. For the Literacy Environment Checklist, the weighted kappa was .84, and for the Group/Family Observation, the weighted kappa was .54 (Neuman, Koh, & Dwyer, 2008).

Internal Consistency

The Literacy Environment Checklist showed good internal consistency ($\alpha = .82$). Cronbach's alphas for the individual subscales of the checklist (books, writing, and resources) ranged from .42 to .78.

The Group/Family Observation showed very strong internal consistency ($\alpha = .97$). Cronbach's alphas for the individual subscales of the Group/Family Observation (physical environment for learning, support for learning, adult teaching strategies) ranged from .91 to .94 (Neuman et al., 2007).

Internal Correlations

The major subscales of the CHELLO Literacy Environment Checklist and Group/Family Observation were moderately to highly correlated with each other, indicating that the physical and psychological aspects of the environment were highly related to one another. Correlations ranged from .34 to .97 (Neuman et al., 2007).

A subsequent study assessing the psychometric properties of the CHELLO showed that, total scores for the Literacy Environment were significantly correlated with each summary score on the Observation ($r = .67$, $r = .33$, and $r = .47$, respectively for the Physical Environment for Learning, Support for Learning, and Teaching Strategies). Total scores for the Literacy Environment and the Group/Family Observation were correlated ($r = .52$). This moderate correlation provides support for the fact that the two tools, while complementary, measured somewhat different aspects of the environment, and should be examined separately" (Neuman et al., 2008).

Validity Information

Concurrent Validity

The CHELLO total score (measured in spring 2006) correlated significantly with children's language growth (as measured by the PPVT ($r = .36$, $p < .01$)), phonological skills (as measured by the PALS nursery rhyme ($r = .25$, $p < .05$)), and ability to do language-oriented math problems (as measured by the Woodcock-Johnson Applied Problems test ($r = .28$, $p < .05$)). The CHELLO was not related to children's developing alphabet skills (as measured by the Woodcock-Johnson Letter Identification subtest). Since the CHELLO was not designed to measure this skill, it is not surprising that there was no correlation between the measures.

References and Additional Resources

Neuman, S., Dwyer, J., & Koh, S. (2007). *Child/Home Early Language & Literacy Observation Tool (CHELLO)*. Baltimore, MD: Brookes Publishing.

Neuman, S. B., Koh, S., & Dwyer, J. (2008). CHELLO: The Child/Home Environmental Language and Literacy Observation. *Early Childhood Research Quarterly*, 23, 159-172.

Smith, M. W., & Dickinson, D. K. (2002). *The Early Language and Literacy Classroom Observation*. Baltimore, MD: Brookes Publishing.

Arnett Caregiver Interaction Scale (CIS)

I. Background Information

Author/Source

Source: Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10, 541-522.

(Note that this article does not contain a list of the items on the scale. However, this is the article that is typically cited when the CIS is used.)

Publisher: A copy of the scale can be found in Jaeger and Funk (2001):
Jaeger, E. & Funk, S. (2001). *The Philadelphia Child Care Quality Study: An Examination of Quality in Selected Early Education and Care Settings*. Philadelphia, PA: Saint Joseph's University.

Purpose of Measure

The purpose of this measure is "to rate the emotional tone, discipline style, and responsiveness of teachers and caregivers in a classroom. The items focus on the emotional tone and responsiveness of the caregiver's interactions with children. The scale does not address issues of curriculum or other classroom management issues (such as grouping or flow of activities)" (U.S. Department of Education, 1997, p. 78).

Population Measure Developed With

"Items were developed during pilot observations in Head Start centers in the Charlottesville, Virginia area. . ." (Arnett, 1989, p. 546).

Age Range/Setting Intended For

This measure may be used in early childhood programs.

Ways in which Measure Addresses Diversity

Information not available.

Key Constructs & Scoring of Measure

The Caregiver Interaction Scale (CIS) consists of 26 items usually divided into 4 subscales. Researchers have conducted factor analyses on the 26 items and have found different subscales (e.g., Whitebook et al., 1989).

Observers are asked to rate the extent to which 26 items are characteristic of the child care provider whom they are observing. Items are scored on a 4-point scale from (1) Not at all characteristic to (4) Very much characteristic of the child care provider. The measure usually contains the following subscales:

- *Sensitivity* (10 items)
- *Harshness* (8 items)
- *Detachment* (4 items)
- *Permissiveness* (4 items)

II. Administration of Measure

Who Administers Measure/Training Required

Training Required: You must achieve a .70 inter-rater reliability for two consecutive visits to be a certified Arnett Caregiver Interaction Scale observer (Jaeger & Funk, 2001).

Setting

The CIS is administered in a classroom or family child care home.

Time Needed and Cost

Time: Caregivers should be observed for 45 minutes or more.

Cost: None

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Jaeger and Funk (2001) reported inter-rater reliability coefficients ranging from .75 to .97 between a certified observer and trainees.

Internal Consistency

Cronbach's alphas from the Observational Study of Early Childhood Programs (Layzer et al., 1993):

Warmth/responsiveness (10) = .91

Harshness (7) = .90

Jaeger and Funk (2001) reported coefficients of .81 and higher for the sensitivity (positive interaction), punitiveness, and detachment subscales.

Validity Information

Concurrent Validity

Layzer et al. (1993) found correlation coefficients of .43 to .67 between the CIS and several other measures of child care quality (i.e., Early Childhood Environment Rating Scale (ECERS), Assessment Profile for Early Childhood Programs, Description of Preschool Practices). However, the authors did not expect large coefficients because the CIS focuses more narrowly on an aspect of teacher behavior than the other observation measures.

References and Additional Resources

- Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10, 541-522.
- Jaeger, E. & Funk, S. (2001). *The Philadelphia Child Care Quality Study: An Examination of Quality in Selected Early Education and Care Settings. Philadelphia*. Philadelphia, PA: Saint Joseph's University.
- Layzer, J. I. (1993). *Observational Study of Early Childhood Programs. Final Report. Volume I: Life in Preschool*. (ERIC # ED366468). Washington, DC: US Department of Education.
- Love, J. M., Meckstroth, A., & Sprachman, S. (1997). *Measuring the quality of program environments in Head Start and other early childhood programs: A review and recommendations for future research* (Working Paper No. 97-36). Washington, DC: U.S. Department of Education National Center for Education Statistics.
- Whitebook, M., Howes, C., & Phillips, D. (1989). *Who cares? Child care teachers and the quality of care in America. Executive summary of the National Child Care Staffing Study*. Oakland, CA: Child Care Employee Project.

Classroom Assessment Scoring System (CLASS)

I. Background Information

Author/Source

Source:

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System Manual K-3*. Baltimore, MD: Brookes Publishing.

Publisher:

Paul H. Brookes Publishing Co.
Post Office Box 10624
Baltimore, MD 21285-0624
Phone: 800-638-3775
Website: 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 third grade 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.

- 30 toddler classrooms, ages 15-36 months (Thomason & La Paro, 2009)
- 694 pre-school classrooms in 11 states; 730 kindergartens in 6 states (National Center for Early Development and Learning MS and SWEEP studies)
- 164 pre-school classrooms in Virginia (MyTeachingPartner Study)
- 82 3rd – 5th grade classrooms in New York City (4R's Study)
- 88 1st – 5th grade 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 1st, 3rd, and 5th grades 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 pre-school through third grade classrooms. Currently two versions of the CLASS are available: a pre-school 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).

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., 2008; Mashburn et al., 2008; Pianta et al., 2005).

Key Constructs & 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 pre-school to third grade period. Each dimension is rated on a 7-point Likert-type scale. The Pre-K and K-3 manuals describe 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
 - Literacy Focus (Not a part of published manual but available upon request)

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 & 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.

II. 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 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 2 hours total) in order to get an accurate sampling of classroom quality data across the three CLASS domains. Total time will vary dependent on the purpose of the observation.

Cost: Two-day training at UVA: \$670/person
Five-day training (Train the trainer): \$1,500/person
Local Training: \$3,000 for up to 15 people (plus travel costs for 1 trainer)

Pre-K Manual: \$49.95

K-3 Manual: \$49.95

Pack of 10 scoring forms: \$25

III. 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. 95-96) as 87%.

Two observers both coded a total of 33 30-minute digital videotapes submitted by teachers in the MyTeachingPartner (MTP) Study. Inter-rater reliability (within 1 point of each other) ranged from 78.8% (for Behavior Management and Instructional Learning Formats) to 96.9% (for Productivity). Similar levels of reliability have been obtained in live observations (Hamre et al., 2008, p. 96).

Internal Consistency

Correlations among the CLASS dimensions range from .11 to .79. Correlations for the pre-school sample in the MS/SWEEP Studies were generally lower than those for the third grade 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., 2008). 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 consistency 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 pre-school and 3rd grade classrooms using data from the NCEDL MS Study of pre-school and the 4R's Study of 3rd grade classrooms in New York City. For the 3rd grade 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 pre-school sample, correlations between the first 4 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 pre-school and 3rd grade (Hamre et al., 2008, p. 97).

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., 2008, p. 99).

CLASS scores have also been found to be relatively stable across the school year, at least in a large number of pre-school classrooms. Analyses also indicate that 7-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 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 Environmental Rating Scale (ECERS-R) classrooms with higher CLASS scores were rated higher on interactions factor (correlations range from .45 to .63) from the ECERS. 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 pre-school, as well as gains in their performance across the pre-school year (Howes et al., 2008; Mashburn 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 domain for predicting achievement was the Instructional Support of the classroom as assessed by the CLASS; the CLASS Emotional Support domain was associated with growth in children's expressive and receptive language scores, increased social competence, as well as decreases in teacher-reported behavior problems (Howes et al., 2008; Mashburn et al., 2008). In addition, the Classroom Organization domain has been linked to children's self-control, engagement, and literacy gains (Ponitz, Rimm-Kaufman, Brock, & Nathanson, 2009; Rimm-Kaufman, Curby, Grimm, Nathanson, & Brock, 2009).

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 and Additional Resources

- Hamre, B. K., Mashburn, A. J., Pianta, R. C., LoCasale-Crouch, J., & La Paro, K. M. (2006). *Classroom Assessment Scoring System Technical Appendix*.
- Hamre, B. K., & 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., & Barbarin, O. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly*.
- La Paro, K. M., & 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., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the prekindergarten year. *The Elementary School Journal, 104*(5), 409-426.
- Mashburn, A. J., Pianta, R., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., Burchinal, M., Clifford, R., Early, D., Howes, C. (2008). Measures of Classroom Quality in Pre-Kindergarten and Children's Development of Academic, Language and Social Skills. *Child Development, 79*, 732-749.
- Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & 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., & Hamre, B. K. (2008a). *Classroom Assessment Scoring System Manual Pre-K*. Baltimore, MD: Brookes Publishing.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008b). *Classroom Assessment Scoring System Manual K-3*. Baltimore, MD: Brookes Publishing.
- Ponitz, C. C., Rimm-Kaufman, S. E., Brock, L. L., & Nathanson, L. (2009). Early adjustment, gender differences, and classroom organizational climate in first grade. *The Elementary School Journal, 110*(2), 142-162.

Rimm-Kaufman, S. E., Curby, T. W., Grimm, K., Nathanson, L., & Brock, L. L. (2009). The contribution of children's self-regulation and classroom quality to children's adaptive behaviors in the kindergarten classroom. *Developmental Psychology, 45*, 958-972.

Thomason, A. C. & La Paro, K. M. (2009). Measuring the quality of teacher-child interactions in toddler child care. *Early Education and Development, 20*, 285-304.

Classroom Assessment Scoring System: Toddler Version (CLASS Toddler)

I. Background Information

Author/Source

Source: Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2009). *Classroom Assessment Scoring System: Toddler version (CLASS Toddler)*. Draft manuscript.

Publisher: This measure is currently unpublished.

Purpose of Measure

The Classroom Assessment Scoring System: Toddler Version (CLASS Toddler) is an observational instrument developed to assess classroom quality in toddler child care classrooms. Similar in format to the CLASS Pre-K and CLASS Elementary (Pianta, La Paro, & Hamre, 2008), the Toddler version captures the average experience of a child in a given classroom, paying particular attention to the teachers' interactions and behaviors with the children. Other measures of classroom quality focus primarily on the physical aspects of the environment or on characteristics of teacher sensitivity and emotional support; however, they do not address important aspects teacher-child interactions or "the 'how' in teaching behaviors," such as behavior guidance and the facilitation of language (Thomason & La Paro, 2009, p. 288). The CLASS Toddler was designed to capture these aspects of process quality specifically in toddler classrooms.

Population Measure Developed With

The CLASS Toddler was piloted with 46 toddler teachers in 30 different toddler classrooms located within a large county in a mid-size southeastern state in the United States. One classroom from each center was chosen based on the following criteria: (1) the classroom was part of a center-based child care center; (2) all of the children in the classroom were between 15 and 36 months of age; and (3) the teacher(s) had been employed by the facility and working in their classroom(s) for at least 1 month. Classrooms had between 3 and 15 children ($M = 8.85$) and 1 to 6 teachers ($M = 2.12$). All of the teachers were female. Over half of the teachers (54%) were Caucasian ($n = 24$); 39% were African American ($n = 17$), 5% were biracial ($n = 2$), 2% reported "Other" as their ethnicity ($n = 1$); and 5% did not report their race/ethnicity ($n = 2$). The mean age of the participating teachers was 34 years, with a range of 18 to 70 years. Teachers had an average of 9 years of experience in child care, with a range of 0 to 35 years. Participating teachers varied widely in their reported highest educational level: 2% completed some graduate work or a Master's degree; 19% had a Bachelor's degree; 9% had an Associate's degree; 42% reported having some college courses but no degree; 2% reported having other educational experience; 21% were high school graduates; and 5% had less than a high school diploma. Seventy-three percent of teachers reported having had training specific to

working with toddlers within the past year and 19% reported belonging to an early childhood professional organization (Thomason & La Paro, 2009).

Age Range/Setting Intended For

The CLASS Toddler is designed to be used in toddler child care classrooms with children between the ages of 15 and 36 months.

Ways in which Measure Addresses Diversity

The instrument is being tested with a racially diverse sample of teachers with various years of experience and educational backgrounds.

Key Constructs & Scoring of Measure

The CLASS Toddler addresses 9 key dimensions:

- *Positive Climate*
- *Negative Climate*
- *Teacher Sensitivity*
- *Regard for Child Perspectives*
- *Behavior Guidance*
- *Facilitation of Classroom Routine*
- *Facilitation of Learning and Development*
- *Quality of Feedback*
- *Language Modeling*

The overall definition of each dimension has remained similar to the pre-k CLASS, however the indicators have been modified to represent the developmental level of toddlers.

Each dimension is scored individually on a 7-point Likert-type scale based on the observation of specific behavioral markers described in the user's manual (e.g., physical proximity; expanding children's involvement). The CLASS is an inferential measure and not a checklist. Observers view the dimensions as holistic descriptions of the average child's experiences in the classroom that fall in the "low" (1, 2), "mid" (3, 4, 5), and "high" (6, 7) range. To complete the ratings, observers must make judgments based on the range, frequency, intention, and tone of the interpersonal and individual behavior observed during each observation cycle. They weigh all behavioral markers equally in each dimension in order to determine the range and assign a final score; however, not all markers or indicators of a particular dimension must be seen in order to score in that range.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Trained CLASS observers administer the instrument across a series of 30-minute observation cycles. Each cycle of observation consists of a 20-minute period during which the observer watches classroom interactions and takes notes, followed by a 10-minute period for scoring. Cycles continue without

interruption until the end of the observation period, with a minimum of four cycles recommended for reliability of measurement.

Training Required: Extensive training and a reliability check are required to appropriately use the CLASS. Scheduled training sessions given by the developers are not yet available for the Toddler version of the instrument.

Setting

The CLASS Toddler is appropriate for use in toddler center-based classrooms with children ages 15 to 36 months.

Time Needed and Cost

Time: A minimum of two hours is required to accurately complete the CLASS Toddler. Observations start at the time the classroom day begins, or at another predetermined time, and continue throughout the morning session. Observers should discuss with the teacher the schedule for the day and use that information to plan accordingly to maximize the length of the observation period and the number of cycles that can be obtained.

Cost: This measure is currently unpublished.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Inter-rater reliability was established through the use of videotaped classroom observations from classrooms not participating in the pilot study. Each observer viewed 5 videotaped segments and independently coded each segment on 6 dimensions (Note: Facilitation of Learning and Development was added after the pilot study). Inter-rater reliability was established at 80% within 1 point on the scale across the 5 videotaped segments (Thomason & La Paro, 2009).

Internal Consistency

A Cronbach's alpha of .88 was obtained on the four dimensions relating to emotional climate—Positive Climate, Negative Climate, Teacher Sensitivity, and Regard for Child Perspectives—indicating adequate internal consistency in that domain (Thomason & La Paro, 2009).

Validity Information

Construct Validity

"Construct validity was established through careful and thorough reviews of existing measures, a review of the research on the unique aspects of toddler development, and observations conducted in toddler child care classrooms. Existing instruments reviewed for this purpose included the ITERS (Harms et al., 2003), the CIS (Arnett, 1989), and the ORCE used in the NICHD Study of Early Child Care (see NICHD ECCRN, 1996). The examples and indicators included in the adapted [CLASS]

measure reflect this review... Additionally, the measure was reviewed by an infant/toddler expert to ensure the validity of the adapted constructs." (Thomason & La Paro, 2009, p. 295). "Validity of the adapted measure was further supported by correlational data between results on the measure and traditional correlates of quality in early childhood education, including teacher-child ratio, group size, teacher education, and child care quality rating systems" (Thomason & LaParo, 2009, p. 297).

Additional validity information will be forthcoming after additional data from pilot sites is collected and analyzed.

Comments

A separate version of the CLASS will be developed for infant child care classrooms.

References and Additional Resources

- Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology, 10*, 541–552.
- Harms, T., Cryer, D., & Clifford, R. M. (2003). *Infant/Toddler Environment Rating Scale, revised edition*. New York: Teachers College Press.
- National Institute of Child Health and Human Development Early Child Care Research Network. (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly, 11*, 269–306.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2009). *Classroom Assessment Scoring System: Pre-K version (CLASS)*. Baltimore: Brookes.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2009). *Classroom Assessment Scoring System: Toddler version (CLASS Toddler)*. Draft manuscript.
- Thomason, A. C., & La Paro, K. M. (2009). Measuring the quality of teacher-child interactions in toddler child care. *Early Education and Development, 20*(2), 285–304.

Classroom CIRCLE: Classroom Code for Interactive Recording of Children's Learning Environments (CIRCLE)

I. Background

Author/Source

Source: Atwater, J., Lee, Y., Montagna, D., Reynolds, L., & Tapia, Y. (2009). *Classroom CIRCLE: Classroom Code for Interactive Recording of Children's Learning Environments*. Kansas City, KS: Juniper Gardens Children's Project.

Publisher: Juniper Gardens Children's Project
650 Minnesota Avenue, 2nd Floor
Kansas City, KS 66101

Purpose of Measure

As described by the authors

The Classroom Code for Interactive Recording of Children's Learning Environments (Classroom CIRCLE), a computerized observation system, was developed to provide a detailed assessment of young children's classroom environments, including: (a) the context of children's classroom activities, (b) the behavior of teachers and other adults in the classroom, and (c) the child's engagement with people and objects. Data are collected on a time-sampling basis using a PDA. The first version of Classroom CIRCLE was developed for a national evaluation of early childhood programs in the Bounce Educare Network. The current revision, with an enhanced focus on early language and literacy, is being used by Early Reading First programs and by the Center for Response to Intervention in Early Childhood.

Population Measure Developed With

Classroom CIRCLE has been developed through observation of a variety of early childhood programs for pre-school children, including Head Start classrooms and community-based child care centers. These programs have included children who are English language learners, as well as children who are at risk for developmental delay.

Age Range/Setting Intended For

While an earlier home-based CIRCLE system was designed for children from 6 to 72 months of age, Classroom CIRCLE is most appropriate for pre-school environments.

Ways in Which Measure Addresses Diversity

Observers note the primary language for each child being observed. Observers also note whether the child uses conventional words in a language other than English or uses sign language.



Key Constructs & Scoring of Measure

This measure consists of 8 variables. A data program was developed to permit data collection with PDA computers. Observers are signaled with a tone at the beginning of the observation interval. The observer selects a category within each variable that best describes the behaviors and events that occurred at the moment of the tone. The categories within each variable are exhaustive and mutually exclusive. Variables include:

- *Context*: Describes typical activities that occur across a child's day and that serve as the context for children's interactions with the environment. This is recorded at the beginning of each 10-minute observation period. If the context changes, the observer records the change.
 - Centers
 - Story Time
 - Large Group Activity
 - Small Group Activity
 - Individual Activity
 - Meals and Snacks
 - Clean-Up, Set-Up, Transition
 - Personal Care
 - Therapy
 - Restricted Access
 - None of Those Listed
- *Teacher Variables*: There are four teacher variables including: Verbal Response, Recipient of Verbal Response, Involvement, and Focus of Instruction. These variables describe the Teacher's behavior toward the Focus Child.
 - Verbal Response: Categorizes Teacher's verbal or vocal behavior toward the Child or the Child's group.
 - Positive Feedback
 - Negative Feedback
 - Expansion, Repetition, Extension
 - Open-Ended Request for Communication
 - Closed-Ended Request for Communication
 - Request for Action
 - Reading, Reciting
 - Singing
 - Exuberant Vocalization, Laughter
 - General Conversation
 - None
 - Recipient of Verbal Response: Indicates the children receiving the Teacher's verbal response
 - Focus Child Only
 - Child's Group
 - None



- Teacher Involvement: Describes the extent of the Teacher's engagement in the Child's activity
 - Sharing Child's Activity
 - Close Proximity
 - General Supervision
 - Not Involved
- Focus of Instruction: Describes teacher strategies for supporting the Focus Child's development in language and early literacy
 - Phonological Awareness
 - Alphabet/Print Concepts
 - Comprehension – Story
 - Comprehension – Other
 - Vocabulary
 - Reading
 - None of Those Listed
- *Child Variables*: There are three Child variables including: Communication and Social Behavior, Social Partner, and Engagement. These variables are used to describe the behavior of the Focus Child.
 - Communication and Social Behavior: Describes the Child's behavior with other people in the classroom.
 - Negative Social Behavior
 - Words – English
 - Words – Other Language, Blends, Signs
 - Communicative Gesture, Vocalization
 - Nonverbal Positive Social Initiation
 - Singing
 - Laughing
 - Social Attention
 - None
 - Social Partner: Observer selects a category to identify the recipient of Social Behavior when Social Behavior has been recorded
 - Teacher
 - Other Professional
 - Other Adult
 - Individual Child
 - Group
 - No Social Partner
 - Engagement: Describes Child's participation in classroom activities
 - Competing Behavior
 - Writing
 - Reading Words or Letters Aloud
 - Other Academic Manipulation
 - Other Academic Verbal Response
 - Academic Attention
 - Music and Recitation
 - Pretend Play



- Non-Academic Manipulation
- Gross Motor Behavior
- Eating and Drinking
- Non-Academic Attention to Materials
- None of Those Listed

Scoring: The Classroom CIRCLE software includes modules for calculating percentage scores and conditional probabilities of occurrence (e.g., the probability of selected target behaviors given selected antecedent conditions). The software produces reports based on single observations, as well as files for archiving data collected across multiple observations.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Observations are conducted by trained observers who are not participating in classroom activities.

Training Required: Classroom CIRCLE materials include a detailed manual and practice exercises. To establish reliability as an observer, one must be trained by a person who has met standards for reliable and accurate administration. With preliminary study and practice, most observers can achieve these standards with 2-4 days of in-person training. The CIRCLE software includes modules for calculating percentage agreement and Cohen's kappa. Future plans include videos and on-line resources to support training.

Setting

The Classroom CIRCLE is designed for use in for pre-school environments.

Time Needed and Cost

Time Needed: One observation session usually includes 3 Focus Children and lasts about 1.5 hours. The observer focuses on one child at a time, switching to a different focal child every 10 minutes, until 30 minutes of data are collected for each child. During each 10-minute segment, data are recorded during a series of 40 15-second intervals. The observer alternates between: (a) 15-second intervals for recording teacher variables, and (b) 15-second intervals for recording child variables.

Cost: Please contact Dr. Jane Atwater for more information (janea@ku.edu).

III. Functioning of Measure

Reliability Information

During 2009 and 2010, research is planned that will provide psychometric information about the current revision.



Inter-rater Reliability

In previous studies using the original CIRCLE, inter-rater reliability averaged 92.0%, and a .75 Kappa value (Atwater, 1999).

Internal Consistency

Child and parent variables assessed with the CIRCLE have been found to be significantly correlated with young children's cumulative risk status (e.g., Atwater & Williams, 1996), with indicators of developmental resilience in infancy (Atwater, 1999), and with pre-school children's early literacy skills (Rush, 1999).

Validity Information

There have been not tests of validity at this time.

References and Additional Resources

Atwater, J., Lee, Y., Montagna, D., Reynolds, L., & Tapia, Y. (2009). *Classroom CIRCLE: Classroom Code for Interactive Recording of Children's Learning Environments*. Kansas City, KS: Juniper Gardens Children's Project.

Atwater, J. (1999, April). Naturalistic assessment of parent-child interaction for families in home-based intervention. In S. McBride (Chair), *Interaction processes in Early Head Start intervention studies: What we can learn from observational studies*. Symposium conducted at the biennial meeting of the Society for Research in Child Development, Albuquerque, NM.

Atwater, J., & Williams, R. (1996, June). *Describing child behavior and child-caregiver interactions in families at risk*. Presented at the Head Start National Research Conference, Washington, DC.

Rush, K. L. (1999). Caregiver-child interactions and early literacy development of preschool children from low-income environments. *Topics in Early Childhood Special Education, 19*, 3-14.

Classroom Language and Literacy Environment Observation (CLEO)

I. Background Information

Author/Source

Source: Holland Coviello, R. (2005). *Language and literacy environment quality in early childhood classrooms: Exploration of measurement strategies and relations with children's development*. State College, PA: Pennsylvania State University.

Publisher: Dissertation published by:
ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Purpose of Measure

As described by the author:

The various elements of the CLEO are meant to address several elements of pre-school language and literacy classroom environments that research has shown are important for affecting children's learning, including the quantity and quality of teacher language input, language and literacy teaching, and children's access to literacy materials in the classroom.

Population Measure Developed With

The CLEO was developed with 16 urban pre-kindergarten classrooms that served Head-Start eligible children and 20 mostly rural Head Start classrooms.

Age Range/Setting Intended For

The CLEO is intended for classrooms that serve children ages 3-5, including pre-school, pre-K, center-based child care, and Head Start settings.

Ways in which Measure Addresses Diversity

One of the programs in which the CLEO was developed served a number of families for whom English was not a first language. The Language Interaction Ratings were thus written to include descriptors that could indicate quality of interaction between a teacher and a child learning English. These descriptors were to be considered when making ratings in addition to the other scale descriptors only when a child or children for whom English was not a home language were present in the classroom.

Key Constructs & Scoring of Measure

There are 5 major elements to the CLEO.

- *Literacy Environment Inventory (CLEO-LEI)*: modified and expanded version of the Early Language and Literacy Classroom Observation's Literacy Environment Checklist (ELLCO LEC; Smith, Dickinson, & Sangeorge, 2002). This section is meant to assess the structural elements of the classroom's literacy environment, such as the presence and availability of books and writing supplies.
- *Language Interaction Observation (CLEO-LIO)*: a simplified version of the Teacher-Child Verbal Interaction Profile (TCVI; Dickinson, Haine, & Howard, 1996) coding scheme, with new coding categories that are theoretically derived from the language development literature. The categories include brief and extended comments, open-ended questions, closed-ended questions, and directives, as well as decontextualized talk.
- *Language Interaction Ratings (CLEO-LIR)*: addresses the sensitivity and cognitive challenge of teachers' verbalizations to children. These items are rated on a 1-5 scale.
- *Literacy Activities Inventory (CLEO-LAI)*: adapted from the ELLCO's Literacy Activities Rating Scale (LARS). New items on bookreading and writing in this section expand upon the ELLCO LARS. Some items were rewritten to focus on teacher behaviors.
- *Literacy Activities Rating scale (CLEO-LAR)*: similar to the CLEO-LIR section, this group of items is rated on a 5-point scale. The items assess the extent to which literacy activities, interaction, and instruction observed in the classroom is developmentally appropriate, and integrated into the social environment of the classroom.

Comments

A CLEO observation should last throughout the classroom day so that all relevant elements are observed. CLEO-LIO coding and LIR ratings can be completed during different classroom activities, such as mealtime, free play, and bookreading, to capture a variety of patterns of language use (see Gest, Holland-Coviello, Welsh, Eicher-Catt, & Gill, 2006). The remaining portions of the CLEO can be completed at the end of the observation, though it is advised that observers keep counts and notes of activities throughout the observations.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The CLEO is most successful when administered by people who have at least bachelors degrees plus experience teaching and/or observing in early childhood classrooms.

Training Required: In the original study, observers participated in a one-day training session. Master-coded video-tapes were used to establish reliability for the LIO and LIR language coding. Participants also participated in live classroom observations to establish reliability on verbalization coding, rating scales, and checklists.

Setting

This measure is intended for use in formal child care settings, pre-school or pre-K classrooms, and Head Start classrooms.

Time Needed and Cost

Time: For training, sites should allow at least one day for initial training, plus 6-8 weeks of practice and reliability training with video-tapes and live in classrooms to establish inter-rater reliability among observers. Once reliability has been established, independent observers should plan to spend one classroom day in each classroom to be assessed. It is recommended that at least 10% of all observations be done in pairs to ensure ongoing inter-rater reliability.

Cost: The author will provide the measure free of cost. Training can be provided by the author and costs \$40/hour for prep (8 hours) and training time plus travel expenses. The author may be contacted by email at rholland-coviello@air.org.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

In the 36 classrooms in which the CLEO was developed, six classrooms (17%) were observed by two people. Pairs consisted of the author and one of two beta observers. Percent agreement for LEI items averaged 95% and 100% across all items for each beta observer with the alpha observer. Similarly, percent agreement for LAI items averaged 93% and 100% across all items for each beta observer with the alpha observer. Average percent agreement within 1 scale-point for LAR was 94% for each pair of observers. Intraclass correlations were also used to determine the reliability of average total scores on the LAR, and they were .80 and .93 for each pair of observers.

LIO coding and LIR were different since there was more than one of these observations per class. There were 225 language observations in all, and 47 (21%) of those were completed in pairs of the alpha observer with one beta observer or the other. Intraclass correlations were used to ascertain Inter-rater reliability again. Intraclass correlations were computed for each Utterance category-- directives, yes/no questions, open-ended questions, brief answers, extended answers, brief comments, extended comments, decontextualized comments, and pretend comments-- with each pair of observers. As Table 1 reveals, intraclass correlations were adequate, or higher than .60, for all but two Utterance categories: brief answers and pretend utterances.

Both of these were low-frequency codes, which contribute to the low reliability, but solutions were devised to make the codes usable in analyses. Brief answers were collapsed with brief comments, and extended answers with extended comments. The reliability for these collapsed coding categories was sufficient with intraclass correlations ranging from .64 to .99. For pretend comments, the code was

dichotomized so that any coding of pretend in the observation yielded a score of 1 and no coding of pretend utterances yielded a score of 0. Kappa was used to determine the reliability of this less rigorous coding and revealed an adequate agreement of 0.64 and 1.00 for the two pairs of observers. Finally, intraclass correlations between the two observers in each pair for the two subscales of the LIR revealed adequate agreement (ICC's range from 0.67 to 0.92).

Table 1: *Reliability of CLEO Utterance Coding and Language Ratings (Intraclass Correlations)*

CLEO Utterance Category	Alpha, Beta 1	Alpha, Beta 2
Directives	0.85	0.97
Yes/no questions	0.95	0.97
Open-ended questions	0.94	0.98
Brief answers	0.70	0.02
Extended answers	0.81	0.60
Brief comments	0.73	0.60
<i>Brief comments + brief answers</i>	0.74	0.64
Extended comments	0.90	0.99
<i>Extended comments + extended answers</i>	0.90	0.99
Decontextualized comments	0.94	0.97
Pretend comments	0.20	0.47
<i>Kappa: some pretend vs. no pretend</i>	0.64	1.00
CLEO LIR Subscales		
Sensitivity and Responsiveness	.72	.88
Richness of Talk	.67	.92

Internal Consistency

CLEO-LEI: alpha = 0.66
 CLEO-LIR: alpha = 0.81
 CLEO-LAR: alpha = 0.83
 CLEO-LAI: alpha= 0.66

Validity Information

Construct Validity

Strong correlations in expected directions among scores on the different elements of the CLEO demonstrate its construct validity. See Table 2 below for correlations.

Concurrent Validity

There have not yet been enough CLEO data collected to complete concurrent validity analyses.

Convergent & Discriminant Validity

CLEO data have been compared to ECERS and ELLCO data from the same classrooms. The CLEO subscales were generally associated with ELLCO subscales in expected ways, confirming the hypothesis that the CLEO and ELLCO measure similar aspects of classroom environments. The possibility of observer bias does warrant caution in interpreting these results. The author completed the majority of both the ELLCO and CLEO observations. The nature of CLEO Utterance coding is objective relative to rating scales, though, and the classroom-level teacher language use variables were associated with the ELLCO in moderately strong and expected ways. Moreover, the ELLCO was completed at least 1 month before the CLEO observations were begun. CLEO convergence with ECERS-R observations was lower than that with the ELLCO. See Table 2 for correlations between scores on each measure.

	1	2	3	4	5	6	7	8
CLEO								
1. Literacy Materials								
2. Utterances: Classroom Rate of High Quality Talk	.13							
3. Utterances: Classroom Proportion of High Quality Talk	.03	.48**						
4. Utterances: Classroom Rate of Directives	.03	.16	-.56**					
5. Utterances: Classroom Proportion of Directives	-.10	-.58**	-.67**	.64**				
6. Language Ratings: Classroom Sensitivity and Responsiveness	.09	.59**	.26	-.12	-.66**			
7. Literacy Activities	.54**	.17	.04	.12	-.01	-.06		
8. Literacy Ratings	.55**	.41*	.20	-.03	-.43**	.54**	.54**	
ELLCO								
9. Literacy Materials	.73**	.31	.37	-.12	-.30	.11	.40	.56*
10. General Classroom Environment Ratings	.33	.30	.33	-.53*	-.76**	.78**	-.05	.37
11. Language, Literacy, & Curriculum Ratings	.52*	.20	.51^	-.47	-.54*	.53*	.11	.39
12. Literacy Activities	-.12	.36	.14	.09	-.18	.60*	-.17	.16
ECERS-R								
13. Total	.33	.07	-.30	-.06	-.18	-.01	-.20	.11
14. Language-Reasoning	.45	.10	-.03	-.16	-.26	-.14	-.03	.20

Table 2: *Correlations Among CLEO, ELLCO, and ECERS-R Variables*

^ $p < .10$, * $p < .05$, ** $p < .01$

Predictive Validity

Preliminary analyses on a small sample have thus far not confirmed the CLEO's predictive validity. Data continue to be collected, so future analyses may provide evidence of the CLEO's predictive validity.

Content Validity

See Holland Coviello (2005, Chapters 2-3) for a literature review connecting CLEO elements and items with research identifying important aspects of environments for children's language and literacy development.

References and Additional Resources

- Dickinson, D. K., Haine, R. A., & Howard, C. (1996). *Teacher-Child Verbal Interaction Profile*. Newton, MA: Education Development Center, Center for Children and Families.
- Gest, S. D., Holland-Coviello, R., Welsh, J. A., Eicher-Catt, D. L., & Gill, S. (2006). Language development sub-contexts in Head Start classrooms: Distinctive patterns of teacher talk during free play, mealtime and book reading. *Early Education and Development, 17*, 293-315.
- Holland Coviello, R. (2005). *Language and literacy environment quality in early childhood classrooms: Exploration of measurement strategies and relations with children's development*. State College, PA: Pennsylvania State University.
- Smith, M. W., Dickinson, D. K., Sangeorge, A., Anastasopoulos, L. (2002). *Early Language & Literacy Classroom Observation Toolkit: Research Edition*. Baltimore, MD: Paul H. Brookes Publishing.

The Classroom Observation of Early Mathematics Environment and Teaching (COEMET)

I. Background Information

Author/Source

Source: Sarama, J. & Clements, D. H. (2009). *Manual for Classroom Observation of Early Mathematics Environment and Teaching*. University at Buffalo, SUNY.

Publisher: This measure is currently unpublished.

Purpose of Measure

The purpose of this measure is to assess the quality and quantity of mathematics instruction in early education settings. It aims to determine teaching strategies, math content, clarity and correctness of mathematics teaching, and quality of student/teacher interaction (Kilday & Kinsey, 2009).

Population Measure Developed With

Clements and Sarama (2008) developed the COEMET with the purpose of measuring classroom instruction changes following use and implementation of the *Building Blocks* (Clements & Sarama, 2007) curriculum. It should be noted, however, that it is not specifically linked to that or any other curriculum, but rather, it is based upon research-based early childhood mathematics best practices and was designed to measure the quality and quantity of early mathematics instruction in any classroom.

Age Range/Setting Intended For

"The COEMET instrument is intended specifically for use in the early childhood setting" (Kilday and Kinsey, 2009, p. 369). Although an exact age range is not clearly specified in the COEMET manual, Clements and Sarama (2008) indicate that the measure was designed for classrooms from toddlers to 2nd grade. Although the authors have used it from pre-kindergarten to first grade to date, publications of its use to have focused on pre-kindergarten classrooms.

Ways in which Measure Addresses Diversity

The following Likert-type items address the teacher's differentiation of math activities based on children's level of development and learning:

- Item 11: The *mathematical content* was appropriate for the developmental levels of the children in this class.
- Item 13: The *pace* of the activity was appropriate for the developmental levels/needs of the children and the purposes of the activity.
- Item 16: The *teaching strategies* used were appropriate for the development levels/needs of the children and purposes of the activity.

- Item 28: The teacher adapted tasks and discussions to accommodate the range of children's abilities and development.

Key Constructs & Scoring of Measure

The COEMET is divided into two main sections, Classroom Culture (CC), and Specific Math Activities (SMA), each of which includes several sub-sections (see below). Assessors complete the Classroom Culture section once to reflect their entire observation. In contrast, the SMA form is completed for each observed math activity (from 0 to 12 activities). "A math activity is defined as one that is set up and/or conducted intentionally by the teacher involving several interactions with one or more children, or set up or conducted intentionally to develop mathematics knowledge (this would not include, for instance, a single, informal comment). Also, the activity must persist for more than 30 seconds" (Sarama & Clements, 2009, p. 1-2).

- *Classroom Culture* (9 items)
 - Environment and Interaction
 - Personal Attributes of the Teacher
- *Specific Math Activities* (19 items)
 - Mathematical Focus
 - Organization, Teaching Approaches, Interactions
 - Expectations
 - Eliciting Children's Solution Methods
 - Supporting Children's Conceptual Understanding
 - Extending Children's Mathematical Thinking
 - Assessment and Instructional Adjustment

The majority of items are coded on a Likert scale from Strongly Disagree to Strongly Agree. There are a few items that are coded based on the percentage of time observed from 0 – 100%.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Although specific requirements for test administration are not outlined, the authors used the following criteria in selecting observers: "Retired teachers identified by administrators as expert in early childhood mathematics teaching (in New York) and doctoral students and staff (in California) were trained on the COEMET" (Sarama et al., 2008, p. 102).

Training Required: Training requirements are not specified by authors in the COEMET manual, however the following information regarding observer training was given by the authors: We suggest two full days of training. The first consists of studying the instrument, and applying it with others while viewing a training videotape and comparing one's coding to that of experts. Trainees then use the instrument in classrooms and submit their tapes for critique. The second day consists of reviewing the critiques and again training with videotapes. In addition, Sarama et

al. (2008, p. 102) stated, "[Observers] practiced administering the instrument until they reached a satisfactory reliability."

Setting

Early childhood pre-kindergarten or pre-school classrooms, up to 2nd grade classrooms.

Time Needed and Cost

Time: Observation: "Assessors spend no less than a half-day in the classroom, for example, from before the children arrive until the end of the half-day (e.g., until lunch)." (Clements & Sarama, 2009, p. 1)

Cost: Contact Dr. Sarama at jsarama@buffalo.edu, 716-645-1155.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

"Interrater reliability for the COEMET, computed via simultaneous classroom visits by pairs of observers (10% of all observations, with pair memberships rotated), is 88%; 99% of the disagreements were the same polarity (i.e., if one was *agree*, the other was *strongly agree*)" (Clements & Sarama, 2008, p. 461).

Internal Consistency

Coefficient alpha (inter-item correlations) for the two instruments ranged from .95 to .97 in previous research. Rasch model reliability is .96 for the COEMET.

Validity Information

Predictive Validity

Studies show the COEMET is a good predictor (e.g., $r = .50$) of child gain in measured mathematics achievement. Further, the COEMET is a partial mediator of the effects of mathematics interventions (Clements & Sarama, 2008).

Construct Validity

The COEMET was created based on a body of research on the characteristics and teaching strategies of effective teachers of early childhood mathematics (Clarke & Clarke, 2004; Clements & Conference Working Group, 2004; Fraivillig, Murphy, & Fuson, 1999; Galván Carlan, 2000; Galván Carlan & Copley, 2000; Horizon Research Inc., 2001; NAEYC, 1991; Teaching Strategies Inc., 2001). Each item is connected to one or more of these studies; thus, there is intended overlap between the instruments, with each specialized for its purpose.

References and Additional Resources

Clements, D. H., & Sarama, J. (2009). *Manual for Classroom Observation of Early Mathematics*. University of Buffalo.

- Clarke, D. M., & Clarke, B. A. (2004). Mathematics teaching in K-2: Painting a picture of challenging, supportive and effective classrooms. In R. Rubenstein & G. Bright (Eds.), *Perspectives on teaching mathematics: 66th Yearbook*, 67-81. Reston, VA: National Council of Teachers of Mathematics.
- Clements, D. H., & Conference Working Group (2004). Part one: Major themes and recommendations. In D. H. Clements, J. Sarama & A.-M. DiBiase (Eds.), *Engaging young children in mathematics: Standards for early childhood mathematics education*, 1-72. Mahwah, NJ: Lawrence Erlbaum Associates.
- Clements, D. H., & Sarama, J. (2007). *SRA Real Math Building Blocks PreK*. Columbus, OH: SRA/McGraw-Hill.
- Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal*, 45, 443-494.
- Fraivillig, J. L., Murphy, L. A., & Fuson, K. C. (1999). Advancing children's mathematical thinking in Everyday Mathematics classrooms. *Journal for Research in Mathematics Education*, 30, 148-170.
- Galván Carlan, V. (2000). *Development of an instrument to assess the use of developmentally appropriate practices in teaching mathematics in early childhood classrooms*. Unpublished Doctoral dissertation, University of Houston.
- Galván Carlan, V., & Copley, J. V. (2000). *Early childhood mathematics profile observational checklist*. Houston, TX: University of Houston.
- Horizon Research Inc. (2001). *Classroom observation protocol*. Chapel Hill, NC: Horizon Research, Inc.
- Kilday, C. R., & Kinsey, M. B. (2009). An analysis of instruments that measure the quality of mathematics teaching in early childhood. *Early Childhood Education Journal*, 36, 365-372.
- NAEYC (1991). *Early childhood classroom observation—National Academy of Early Childhood Programs (revised ed.)*. Washington, DC: National Association for the Education of Young Children.
- Sarama, J. & Clements, D. H. (2009). *Manual for Classroom Observation of Early Mathematics Environment and Teaching*. University at Buffalo, SUNY.
- Sarama, J., Clements, D. H., Starkey, P., Klein, A., & Wakeley, A. (2008). Scaling up the implementation of a pre-kindergarten mathematics curriculum: Teaching for understanding with trajectories and technologies. *Journal of Research on Educational Effectiveness*, 1, 89-119.

Teaching Strategies Inc. (2001). *A checklist for assessing your program's implementation of the Creative Curriculum for early childhood* (Third ed.). Washington, DC: Author.

Caregiver Observation Form and Scale (COFAS)

I. Background Information

Author/Source

Source: Fiene, R. (1984). *Child Development Program Evaluation Scale and COFAS*. Washington, DC: Children's Services Monitoring Consortium

Purpose of Measure

As described by the authors:

The Caregiver Observation Form and Scale (COFAS) is used to record behaviors of caregivers while interacting with children in a classroom setting.

Population Measure Developed With

The COFAS was developed to complement the Child Development Program Evaluation Scale (CDPES) in order to assess interactions between teachers and children in child care settings. The items included in the COFAS were included after an extensive review of the research literature on the distinguishing characteristics of high quality programs and their teachers.

Age Range/Setting Intended For

The COFAS can be used with any age group of children between infancy and 12 years of age.

Ways in which Measure Addresses Diversity

The COFAS does not directly measure diversity but it is intended to be used with the CDPES which has an item that addresses diversity.

Key Constructs & Scoring of Measure

There are five key constructs of the COFAS: *Language*, *Socio-emotional*, *Motor*, *Cognitive*, and *Caregiving*. Each is described in detail below:

- *Language* (9 items)
 - Speak unsolicited to child
 - Use the child's dialect
 - Respond verbally to child's speech
 - Read or identify pictures to a child
 - Sing or play music with a child
 - Speak slowly and clearly to a child at all times
 - Interrupt or cut off a child's verbalization
 - Scream or yell at children
 - Allow noise level to become too high it is hard to understand children

- *Socio-emotional* (11 items)
 - Give affectionate physical contact to child
 - Make activity suggestion to child
 - Physically punish child
 - Use food as a reinforcement
 - Make fun of or ridicule a child
 - Let other children make fun of or ridicule a child
 - Verbally criticize, scold or threaten a child
 - Isolate a child physically
 - Ignore a child's request
 - Interrupt a child's activity and prevent its completion
 - Leave the child alone
- *Motor* (1 item)
 - Foster development of gross motor skills
- *Cognitive* (4 items)
 - Show impatience or annoyance with child's questions
 - Use terms which are above a child's reasoning ability
 - Deal in abstract concepts without concrete examples
 - Show intolerance with a child's mistakes
- *Caregiving* (4 items)
 - Prepare or serve food for a child
 - Prepare activities or arrange the room
 - Do nothing
 - Talk with other adults

Each caregiver is observed for 10 consecutive two minute periods, with pauses between observations to record. During the pauses, observers record whether or not they observed each behavior listed on the form. The ten responses for each behavior are then summed and multiplied by a weight (either positive or negative) to yield an interaction score for that behavior. Once all behaviors are individually scored, they may be summed to obtain a total interaction score. This score can then be checked against the COFAS scale where a Level 1 indicates "Good" interaction, Level II indicates "Fair," Level III indicates "Poor," and Level IV indicates "Non-optimal."

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The COFAS can be used by state licensing and monitoring staff, researchers, and directors of early care and education programs.

Training Required: Training on the COFAS requires a half day of classroom instruction followed by 3-4 days of on-site inter-rater reliability testing. Individuals who are interested in using the scale should plan on 1 week of training and on-site implementation before using the scale for actual data collection.

Setting

The COFAS is to be administered in the classroom setting.

Time Needed and Cost

Time: Each caregiver is to be observed for 10 consecutive two minute periods, with pauses between observations to record.

Cost: Free.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Inter-rater reliability showed a kappa of .81

Internal Consistency

Cronbach's Alpha was .89 Total Scale

Validity Information

Construct Validity

Construct Validity was assessed by comparing the COFAS with licensing and program quality assessment decisions and ratings ($r = .61; p < .05$).

Concurrent Validity

Concurrent Validity was assessed by comparing the COFAS and the ECERS total scores ($r = .67; p < .01$).

Comments

The COFAS is intended to be used in conjunction with the CDPES. It is an excellent addition to the CDPES in assessing the behaviors of caregivers while interacting with children in a classroom setting.

For additional information regarding the COFAS, please contact:

Richard Fiene, Ph.D., Associate Professor
Human Development and Family Studies
W-311 Olmsted Building
Penn State University - Harrisburg
777 West Harrisburg Pike
Middletown, Pennsylvania 17057
rjf8@psu.edu

References and Additional Resources

Fiene, R. (1984). *Child Development Program Evaluation Scale and COFAS*.
Washington, DC: Children's Services Monitoring Consortium.

Classroom Practices Inventory (CPI)

I. Background Information

Author/Source

Source: Hyson, M. C., Hirsh-Pasek, K., & Rescorla, L. (1990). The Classroom Practices Inventory: An observation instrument based on NAEYC's guidelines for developmentally appropriate practices for 4- and 5-year-old children. *Early Childhood Research Quarterly*, 5, 475-494.

NICHD Study of Early Child Care and Youth Development. (1995). Instructions for completing the classroom observations. In *Phase II Manuals* (p. 1-41).

Purpose of Measure

The Classroom Practices Inventory is a rating scale designed to assess the developmental appropriateness of classroom and curriculum practices, teacher behaviors, children's activities, and teacher-child interactions. "Developmentally appropriate practices emphasize direct experiences, concrete materials, child-initiated activity, social interaction, and adult warmth" (Love, Meckstroth, & Sprachman, 1997, p. 80).

Population Measure Developed With

"The CPI was developed as part of a 2-year study titled 'Academic Environments in Early Childhood: Challenge or Pressure?' (Hyson, Hirsh-Pasek, & Rescorla, 1989) . . . The CPI was used in the Academic Environments study to observe ten early childhood programs reputed to represent a variety of educational practices. These programs had been selected for the study because they had reputations in the community as being either relatively academic or relatively unstructured and play oriented. Located in Pennsylvania and Delaware, these half-day private pre-schools served middle and upper middle class families. All were 4-year-old or pre-kindergarten programs.

The sample was supplemented by including observations of 48 additional programs by university students in early childhood education courses. These programs represented a wider range of settings, including half-day pre-schools, laboratory schools, day care centers, and public and private kindergartens in Pennsylvania and Delaware.

In all, the CPI was used in 207 separate observations of 58 early childhood programs, with a mean of 3.5 observations of 58 early childhood programs" (Hyson et al., 1990, p. 479).

Age Range/Setting Intended For

The CPI is intended to be used in early childhood programs for 4- and 5-year-old children.

Ways in which measure addresses diversity

The NAEYC guidelines incorporate cultural and linguistic diversity as an element of developmentally appropriate practices. In the manual developed for the NICHD Study of Child Care, Item #1-3 asks the observer to take note of "any evidence of cultural awareness as indicated in artifacts in the classroom: holiday displays, pictures, picture books, posters, etc." (NICHD, 1995, p. 41).

The NICHD observer manual also explains that the observer should look for any dolls or toys that represent different ethnic groups while also looking for evidence of ethnic or religious traditions. It is stressed that the observer should look for the representation of multiple ethnicities and religions.

Key Constructs & Scoring of Measure

The CPI is a 26-item rating scale, based on the 1985 edition of the NAEYC guidelines for developmentally appropriate practices. Each item is rated on a 5-point Likert-type scale from 'not at all like this classroom' to 'very much like this classroom.'

- *Developmentally appropriate practices* (10 items)
- *Developmentally inappropriate practices* (10 items)
- *Emotional climate* (6 items). —These items tap teachers' warmth, encouragement, and positive guidance, as well as the overall affective tone of the classroom" (Hyson et al., 1990, p. 478-479).

The measure was developed before NAEYC's 1997 revision of the Developmentally Appropriate Practice position statement and guidelines (Bredekamp & Copple, 1997). As a result, the constructs used in the measure do not reflect revisions that placed more importance on (a) a broader range of teaching strategies, (b) cultural and individual adaptations of classroom practices, and (c) the place of academic content within an "appropriate" early childhood classroom environment.

II. Administration of Measure

Who Administers Measure/Training Required

Administration of Measure: —Ratings are based on several hours of direct observation. In the Academic Environments study (Hyson, Hirsh-Pasek, & Rescorla 1989), 10 programs were visited twice within two weeks by observers with training and experience in early childhood. In addition, 48 day care settings were visited by students in early childhood courses; each program was observed for two and a half hours" (Love et al., 1997, p. 80).

Training Required: "Training of student observers consisted of reviewing complete NAEYC guidelines, reviewing the items, and doing practice classroom observations" (Love et al., 1997, p. 80).

Several weeks of preliminary observations are required for observer training.

Setting

The CPI is appropriate for early childhood programs for 4- and 5-year olds. The measure was adapted for use in kindergarten-primary programs (Vartuli, 1999).

Time Needed and Cost

Time: Observers must spend at least 2.5 hours in the classroom before completing the CPI.

Cost: Information not available.

III. Functioning of Measure

Reliability information

Inter-rater Reliability

Based on observations of 10 programs, inter-rater agreement averaged 64%. Agreement within 1 scale point was 98%. Total CPI scores correlated .86 across pairs of raters (Hyson et al., 1990).

Internal Consistency

- *Developmentally appropriate practices* = .92
- *Developmentally inappropriate practices* = .93
- *Emotional climate* = .88
- Total appropriateness (26 items) = .96

Intercorrelations Among Items

Appropriate and inappropriate program items are highly correlated ($r = -.82$). Emotional climate is highly correlated with program focus ($r = .81$) (Hyson et al., 1990).

Validity Information

Concurrent Validity

CPI scores were related to programs' community reputations as academic or play-oriented and unstructured and to the self-reported educational attitudes of the program teachers (Hyson et al., 1990).

Content Validity

The CPI was developed specifically to operationalize the 1985 NAEYC Guidelines for Developmentally Appropriate Practice. The wording of items closely paralleled the wording of the Guidelines.

References and Additional Resources

Hyson, M. C., Hirsh-Pasek, K., & Rescorla, L. (1990). The Classroom Practices Inventory: An observation instrument based on NAEYC's guidelines for

- developmentally appropriate practices for 4- and 5-year-old children. *Early Childhood Research Quarterly*, 5, 475-494.
- Hyson, M. C., Hirsh-Pasek, K., & Rescorla, L. (1989). *Academic environments in early childhood: Challenge or pressure?* Final report to the Spencer Foundation.
- Layzer, J. I. (1993). *Observational Study of Early Childhood Programs. Final Report. Volume I: Life in Preschool.* (ERIC # ED366468). Washington, DC: US Department of Education.
- Love, J. M., Meckstroth, A., & Sprachman, S. (1997). *Measuring the quality of program environments in Head Start and other early childhood programs: A review and recommendations for future research* (Working Paper No. 97-36). Washington, DC: U.S. Department of Education National Center for Education Statistics.
- NICHD Study of Early Child Care and Youth Development. (1995). Instructions for completing the classroom observations. In *Phase II Manuals*, 1-41.
- Vartuli, S. (1999). How early childhood teacher beliefs vary across grade level. *Early Childhood Research Quarterly*, 14 (4), 489-514.

The Emergent Academic Snapshot (EAS)

I. Background Information

Author/Source

Source: Ritchie, S., Howes, C., Kraft-Sayre, M. & Weiser, B. (2001). *Emergent Academic Snapshot Scale*. Los Angeles: UCLA. This measure is currently unpublished. Adapted from previous instruments for use in the National Center for Early Development and Learning (NCEDL) Multi-State Study of Pre-Kindergarten and the State-Wide Early Education Programs Study (SWEEP) (Early, Barbarin, Bryant, Burchinal, Chang, Clifford, et al., 2005).

Publisher: Available from Carollee Howes
Department of Education
University of California at Los Angeles
Box 951521
Los Angeles, CA
howes@gseis.ucla.edu

Purpose of Measure

The EAS is a time sampling observation instrument designed to describe children's exposure to instruction and engagement in academic activities as well as to describe activities and adult responsive involvement. The unique contributions of EAS as compared to previous observational instruments are in the teacher engagement of the children and children's engagement with academic activities sections.

Population Measure Developed With

The EAS was developed with children from diverse ethnic/racial and home language backgrounds, enrolled in various child care settings from relative or informal care, to center-based care, and children in pre-kindergarten and kindergarten programs.

Age Range/Setting Intended For

The EAS is used with children 10 months to 8 years. The EAS may be used in home and classroom early care and education settings.

Ways in Which Measure Addresses Diversity

The measure was developed in settings with diverse populations. It does not directly measure diversity.

Key Constructs & Scoring of Measure

The 27 items on the EAS are divided into sections including:

- *Children's activity setting* (Howes & Smith, 1995; Kontos, Howes, Galinsky, & Shinn, 1997; Kontos, Howes, Shinn, & Galinsky, 1995)

- *Adult involvement with the child* (Adult Involvement Scale) (Howes, Phillips, & Whitebook, 1992; Howes & Stewart, 1987)
- *Peer Play Scale* (Howes & Matheson, 1992)
- *Teacher engagement of the children*, including codes for seven kinds of instructional strategies (e.g., didactic, uses home language of child)
- *Children's engagement with academic activities* including codes for 14 specific academic activities (e.g., letter-sound correspondence)

II. Administration of Measure

Who administers Measure/Training Required

The measure is collected during naturalistic observation. Observers must be trained and certified by the authors. Observers should have a BA degree and experience working with children.

Setting

The EAS is designed for use in home and classroom early care and education settings.

Time Needed and Cost

Time Needed: The design of the study determines the time sampling frame. The instrument can be used in either a traditional time-sampled procedure – one child at a time – or as a snapshot. When one child at a time is sampled, at least 3 five-minute samples of 15- to 20-second intervals should be collected across a one- to two-hour period. When used in snapshot fashion, up to 4 children can be sampled in succession.

Each Snapshot observation consists of a 20-second observation period, followed by a 40-second coding period. The first child is observed and coded, then the second, third and fourth. When all four children are observed, the observer starts over with the first child. To be reliable each child's behavior should be sampled 45 to 100 times.

Cost: Information not available.

III. Functioning of Measure

Reliability Information

As an observational measure there must be strict standards of reliability with the gold standard, re-establishing reliability every 10th observation, and correction for drift. In the NCEDL and SWEEP studies observer reliability mean weighted Kappa met or exceeded .75.

Validity Information

Concurrent Validity

Teacher engagement of the children and children's engagement with academic activities have modest and positive associations with *the Early Childhood*

Environment Rating Scale-Revised (Early et al., 2005; Harms, Clifford, & Cryer, 1998; Pianta et al., 2005).

Predictive Validity

Children's engagement in academic activities and child assessments in language and literacy were positively associated in fall and spring of pre-kindergarten (Howes et al., in press).

References and Additional Resources

- Early, D., Barbarin, O., Bryant, B., Burchinal, M., Chang, F., Clifford, R., et al. (2005). Pre-kindergarten in eleven states: NCEdL's multi-state study of pre-kindergarten and state-wide early education programs (SWEEP) study. Retrieved December 1, 2005 from http://www.fpg.unc.edu/~ncedl/pdfs/SWEEP_MS_summary_final.pdf
- Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early childhood environment rating scale: Revised edition*. New York: Teachers College Press.
- Howes, C., Burchinal, M., Pianta, R. C., Bryant, D., Early, D., Clifford, R., et al. (2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten. *Early Childhood Research Quarterly, 23*(1), 27-50.
- Howes, C., & Matheson, C. C. (1992). Sequences in the development of competent play with peers social and social pretend play. *Developmental Psychology, 28*, 961-974.
- Howes, C., Phillips, D. A., & Whitebook, M. (1992). Thresholds of quality: Implications for the social development of children in center-based child care. *Child Development, 63*, 449-460.
- Howes, C., & Smith, E. (1995). Child care quality, teacher behavior, children's play activities, emotional security and cognitive activity in child care. *Early Childhood Research Quarterly, 10*, 381-404.
- Howes, C., & Stewart, P. (1987). Child's play with adults, toys, and peers: An examination of family and child-care influences. *Developmental Psychology, 23*, 423-430.
- Kontos, S., Howes, C., Galinsky, E., & Shinn, M. B. (1997). Children's experiences in family child care and relative care as a function of family income and ethnicity. *Merrill Palmer Quarterly, 43*, 386 - 403.
- Kontos, S., Howes, C., Shinn, M., & Galinsky, E. (1995). *Quality in family child care and relative care*. New York: Teacher's College Press.

Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., & 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.

Ritchie, S., Howes, C., Kraft-Sayre, M. & Weiser, B. (2001). *Emergent Academic Snapshot Scale*. Los Angeles: UCLA (Unpublished Instrument).

Early Childhood Classroom Observation Measure (ECCOM)

I. Background Information

Author/Source

Source: Stipek, D. & Byler, P. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly*, 19, 375-397.

Publisher: This measure is currently unpublished. The measure may be obtained by emailing Deborah Stipek at stipek@stanford.edu

Purpose of Measure

As described by the authors:

"Most extant observation measures of early childhood classroom environments focus predominantly on the social climate and resources of the classroom, with less attention given to the quality of instruction provided by the teacher. The Early Childhood Classroom Observation Measure (ECCOM) was developed to tap the nature and quality of academic instruction as well as the social climate, resources, and other aspects of effective classrooms" (Stipek & Byler, undated coding manual, p. 1).

The version of the ECCOM reported on in Stipek and Byler (2004) assesses independently the degree to which constructivist (child-centered) and didactic (teacher-centered) instructional approaches are observed. The measure focuses on the approach used for instruction rather than subject matter content.

The instrument was developed primarily as a research tool. However, at least one research team (Head Start Quality Research Project) is using the ECCOM as an intervention tool as well as for research. "The ECCOM might also be used effectively to help teachers assess and adjust their own practices, or as a tool for principals and directors for assessing teachers" (Stipek & Byler, 2004, p. 392). Thus, the ECCOM may be used for research, as a professional development tool, and/or as a program development and evaluation tool. The value of the ECCOM for professional development purposes has not yet been systematically assessed.

Population Measure Developed With

- 127 kindergarten and first-grade teachers in 99 schools (96 public, 3 private).
- The classrooms represented 46 school districts within 3 states (2 in the northeast, 1 on the west coast).
- Schools were both in urban and rural areas.
- The 127 teachers were predominantly female (n=121) and Caucasian (n=96).
- 234 children were distributed across the classrooms (118 girls, 116 boys; 159 in kindergarten, 75 in first grade).

Age Range/Setting Intended For

This measure is appropriate for classrooms serving children ages 4 to 7, roughly corresponding to the last year of pre-school, kindergarten, and first grade.

Ways in which Measure Addresses Diversity

In the most recent version of the ECCOM there are checklists for "Representations Related to Diversity" and "Treatment of Native Language" (the latter only applies to classrooms in which there are limited or non-English speaking children). However, this is not the version of the measure for which psychometric information is presented.

Key Constructs & Scoring of Measure

The ECCOM reported on in Stipek and Byler (2004) consists of 32 items (17 constructivist, 15 didactic) rated on a scale of 1 (practices are rarely seen) to 5 (practices predominate). There were parallel items for both constructivist and didactic practices, but there were two additional items in the constructivist scale (relevance of instruction activities, and teacher warmth).

The rating of each item occurs after an observation of the classroom. Scores are based roughly on the percentage of time the described practices were seen during observation. Observers are instructed to give a score of 1 if during relevant times the practices described were seen 20% of the time or less; 2 if they were seen 21-40% of the time; 3 if they were seen 41-60% of the time; 4 if they were seen 61-80% of the time; and 5 if they were seen 81-100% of the time. These percentages were used as a guide rather than as an absolute reflection of the frequency of the practices.

- *Constructivist Subscales.*
 - *Instruction.* A high score occurs if children are held accountable for completing work and held to a clear standard, lessons are coherent and well-connected to children's previous knowledge, lessons teach identifiable concepts and are focused on understanding, children are active participants in instructional conversations, and specific strategies for math and literacy instruction are implemented.
 - *Management.* A high score occurs if teachers provide children with choices in both teacher-planned activities and during free time, rules and routines are clear but flexible, children are given developmentally appropriate responsibilities, and discipline is brief and non-disruptive (often involving explanations or assisting children in their own social problem solving).
 - *Social climate.* A high score occurs if teachers are warm, responsive, attentive, and respectful of children.
- *Didactic Subscales.*
 - *Instruction.* A high score occurs if the teacher holds children accountable for completing work and for attaining universal rather than individualized standards, lessons focus on discrete skills, the teacher focuses on facts and procedural knowledge, the teacher

controls the classroom conversations, and math and literacy instruction emphasizes learning distinct skills which are not embedded in meaningful contexts and also strongly emphasizes correctness.

- *Management.* A high score occurs if the rules and routines are teacher-determined, children do not select their own activities outside of recess, and the teacher takes responsibility for maintaining order in the classroom, including intervening quickly in social conflict situations.
- *Social climate.* A high score occurs if there are few social interactions among children, little collaborative work among children, and most children work individually or in a teacher-led group. Tasks and expectations are teacher- or curriculum-driven and uniform across all children.

Comments

The undated Coding Manual (Stipek & Byler) indicates that the most recent ECCOM consists of three parts. Part one is 17 scale items rated on a 1 (practices are rarely seen) to 5 (practices predominate) scale to capture classroom instructional practice. Three types of instructional practice are identified: "best practices" based on a social-constructivist theoretical orientation, teacher-controlled/directed, and child-dominated with little teacher direction or control. The scale items were combined to create six subscales:

- *Social Climate.* The degree to which the classroom climate promotes respect for individuals and individual differences.
- *Learning Climate.* The quality of instruction, coherence of lessons, and standards of learning provided by the teacher.
- *Management.* Child responsibility, choice of activities, and management and discipline strategies employed by the teacher.
- *Math Instruction*
- *Literacy Instruction*
- *Classroom Resources.* The breadth of classroom materials provided for the children in the areas of technology, literacy, mathematics, dramatic play, art, gross motor equipment, and real-life objects.

Part two consists of 10 checklists that assess the instructional and multicultural materials available in the classroom. Part three consists of observers' detailed descriptions of activities and interactions observed over the 3-hour observation period recorded on a Chronology Sheet. Parts two and three should be completed during the observation; part one should be completed at the end of the visit (Stipek & Byler, undated, p. 1).

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Observations are conducted by a trained observer. The authors recommend that observations be conducted on a typical day, and that the observations

begin at the beginning of the day for full-day programs or at the beginning of the program for less-than-full-day programs. Observations occur over a 3-hour period, and should always include observations of math and literacy instruction.

Training Required: Training is required to assure proper use of the instrument. All observers should attend two full days of training and pass a reliability test (i.e., demonstrate 80% reliability on coding with the head trainer or previously certified observer).

Setting

Observations are made in the classroom.

Time Needed and Cost

Time: Three hours are needed for observing the classroom, which includes observing math and literacy instruction.

Cost: Contact Dr. Deborah Stipek at stipek@stanford.edu

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Observers independently rated 26 classrooms in pairs. Intraclass correlations were used to calculate reliability. Reliability was high for all subscales (*Constructivist*: instruction, 0.80; management, 0.92; social climate, 0.82; *Didactic*: instruction, 0.80; management, 0.88; social climate, 0.88; all $p < 0.001$) (Stipek & Byler, 2004, p. 387).

Internal Consistency

Alphas were high for all subscales (*Constructivist*: instruction, 0.73; management, 0.86; social climate, 0.89; *Didactic*: instruction, 0.82; management, 0.87; social climate, 0.91) (Stipek & Byler, 2004, p. 388).

Validity Information

Concurrent Validity

Stipek and Byler (2004) found predictable associations between the ECCOM and teachers' self-reported practices.

Specifically, correlations between the constructivist and didactic subscales of the ECCOM and teachers' self-reported teaching practices revealed the following:

- Teachers who received high scores on the didactic teaching practices scale reported using strategies focused more on basic literacy and math skills ($r = 0.41$ and 0.37 , respectively, $p < 0.001$), were more likely to use skill-based math groups ($r = 0.20$, $p < 0.05$), gave more homework ($r = 0.28$, $p < 0.001$), and planned to retain more children ($r = 0.29$, $p < 0.001$).
- Teachers who received high scores on the constructivist teaching practices scale reported using more inquiry-based math practices ($r = 0.21$, $p < 0.05$),

reported less focus on basic math skills ($r = -0.19, p < 0.05$), expected to retain fewer students ($r = -0.22, p < 0.01$), were less likely to use skill-based math groups ($r = -0.21, p < 0.05$), and gave somewhat less homework ($r = -0.17, p < 0.10$).

Predictive Validity

Correlations between the constructivist and didactic subscales of the ECCOM and teachers' ratings of children's academic skills and self-directed learning revealed the following:

- No relationship between constructivist scores and ratings of children's academic skills and self-directed learning.
- Significant correlation between didactic scores and teachers' ratings of children's math skills ($r = -0.21, p < 0.05$).
- Marginally significant correlation between didactic scores and teachers' ratings of children's self-directed learning ($r = -0.18, p < 0.10$).
- The more didactic the teaching style, the lower teachers rated students on both math skills and self-directed learning.

The authors concede that direct observation of child behaviors and skills would be better than relying on teacher report for assessing associations between the ECCOM and child outcome measures.

Comments

From the coding manual, it is clear that the ECCOM has been updated. Psychometric information on the current version of the ECCOM is needed.

References and Additional Resources

Stipek, D. & Byler, P. (2004). The early childhood classroom observation measure. *Early Childhood Research Quarterly, 19*, 375-397.

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

I. 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, UK
Phone: +44(0) 1782 745567
E-mail: 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, & Taggart, 2003, p. 7).

Population Measure Developed With

"The ECERS-R 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 (3-5 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 pre-school 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, p. 7-8).

Age Range/Setting Intended For

The ECERS-E may be used with children 3 through 5 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 & Scoring of Measure

The ECERS-E supplements the ECERS-R with four new subscales. Items are rated on a 7-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

II. 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 serving children between the ages of 3 and 5, 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 \$475-\$635), and basic training on the ECERS-R and E together

might cost approximately £500-600 (roughly \$795-\$955) per person. Training to research standards (i.e. with appropriate reliability checks) might cost in the region of £1,200 (roughly \$1,910) per person. The scales themselves are priced at £12.99 (roughly \$21) plus delivery costs.¹

III. 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 pre-school settings, they both measure a general construct of 'quality.' Therefore, it is expected that centers obtaining a high score on the ECERS-R will also obtain a high score on the ECERS-E (...)

"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

¹ The conversion rates used reflect rates as of 1/7/10.

moderately associated with all ECERS-E subscales (from .45 to .58)" (Sylva et al., 2003, p. 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 pre-school 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., 2003)" (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 and Additional Resources

- Sammons, P., Sylva, K., Melhuish, E., Siraj-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., & 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.
- 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 Childhood Environment Rating Scale – Revised Edition (ECERS-R)

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

I. Background Information

Author/Source

Source: Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale – Revised Edition*. New York, NY: Teachers College Press.

Harms, T., Clifford, R. M. & Cryer, D. (2005). *Early Childhood Environment Rating Scale – Revised Edition*. New York, NY: Teachers College Press. (Updated with additional notes and a new expanded scoresheet).

Publisher: Teachers College Press
1234 Amsterdam Avenue
New York, NY 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 & Riley, 2003). It can be used by researchers, practitioners, program monitors and early childhood professionals providing 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, & Cryer, 1998, p.1).

A Spanish Language version of the ECERS-R is available from Teachers College Press (www.teacherscollegepress.com/assessment_materials.html). In addition translations of the scale into a variety of other languages are available. Contact the authors (www.fpg.unc.edu/~ecers/) for more information.

Population Measure Developed With

The original ECERS was developed with typical child care programs in North Carolina, but later work and revisions have been based on data from a wide range of program types including Head Start programs, typical child care centers, schools

serving pre-kindergarten and kindergarten children, programs for children with special needs. Special efforts were made to build on input from and experience with programs serving diverse populations including variations in race and ethnicity, type of special need and levels of income. Revisions were based on extensive use of the ECERS in various parts of the US and in other countries.

Age Range/Setting Intended For

The ECERS-R is designed to be used with one room or one group at a time, for children 2 ½ through 5 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 portray positively different races, cultures, ages, gender 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 7-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. It should be noted that indicators under inadequate are scored in the opposite direction from indicators at the higher levels.

- *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
 - Provisions for personal needs of staff
 - Provisions 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.

The ECERS-R and other ERS instruments are also available in electronic form for use on Tablet PC machines through software package developed by the Branagh Information Group (www.ersdata.com) under license from Teachers College Press. This package is most appropriate for medium and large scale users.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The ECERS-R book provides questions 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~~ 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 et al., 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 & Riley, 2003) that offers detailed information and photos that 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 and kindergarten classrooms.

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 & Riley, 2003, p. xiv). The authors recommend that at least 2.5 to 3 hours be spent observing in the classroom and note that spending more than 3 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 \$19.95

Spanish ECERS-R \$19.95

Video Training Packages

1999, VHS \$59.00

2006, DVD \$59.00

Training Workbook

1999, \$4.00

III. 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 et al., 1998, p. 2).

Subsequent use in major studies confirm the reliability when administered by observers trained to reliability and with monitoring of reliability during the data collection period. Care is urged by the authors to avoid conflicts of interest by observers as this has been shown to affect reliability and accuracy.

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 et al., 1998, p. 2). Authors urge care in interpreting the subscale scores.

<i>Space and Furnishings</i>	.76
<i>Personal Care Routines</i>	.72
<i>Language-Reasoning</i>	.83
<i>Activities</i>	.88
<i>Interaction</i>	.86
<i>Program Structure</i>	.77
<i>Parents and Staff</i>	.71
<i>Total</i>	.92

Validity

Concurrent Validity

The Total Score on the ECERS-R has been found to be correlated with two dimensions of the CLASS (Pianta, LaParo, & Hamre, 2008): Emotional Climate, $r = .52$ and Instructional Support, $r = .40$.

The Total Score of the ECERS-R has also been shown to correlate with the ELLCO (Smith, Dickinson, Sangeorge, & Anastasopoulos, 2002) total classroom observation score ($r = .41$) and the Literacy Environment Checklist ($r = .44$).

Finally, the Total Score of the ECERS-R has also been found to be positively correlated with the CIS (Arnett, 1989) ($r = .69$), and the ECERS-E (Sylva, Siraj-Blatchford, & Taggart, 2003) ($r = .78$).

Clifford, Reszka, and Rossbach (2009) also report findings on associations between the subscales of the ECERS-R with several other measures of classroom quality.

Predictive Validity

Math/Number. Research suggests that there is a positive relationship between the social interaction subscale of the ECERS-R and children's early number concept development. The ECERS-R has also been found to be associated with the

Woodcock-Johnson-R (Woodcock & Johnson, 1990) math achievement applied problems subset (Clifford, Reszka, & Rossbach, 2009).

Language/Literacy. Higher scores on the ECERS-R are associated with children's development of receptive language, print awareness, and book knowledge (Clifford, Reszka, & Rossbach, 2009).

Social Outcomes. Researchers have found an association between the total score on the ECERS-R and children's social emotional development (Clifford, Reszka, & Rossbach, 2009). Additionally, several subscales of the ECERS-R have been found to be associated with children's scores on measures of independence, concentration, cooperation, and conformity skills in pre-school (Clifford, Reszka, & Rossbach, 2009).

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 and Additional Resources

- Arnett, J. (1989). Caregivers in day care centers: Does training matter? *Applied Developmental Psychology, 10*, 541-522.
- Clifford, R. M., Reszka, S. S., & Rossbach, H-G. (2009). Reliability and validity of the Early Childhood Environment Rating Scale – Draft version of a working paper. Chapel Hill, NC: University of North Carolina at Chapel Hill, FPG Child Development Institute.
- Cryer, D., Harms, T. & Riley, C. (2003). *All about the ECERS-R: A detailed guide in words & pictures to be used with the ECERS-R*. PACT House Publishing.
- Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale – Revised Edition*. New York, NY: Teachers College Press.
- Harms, T., Clifford, R.M. & Cryer, D. (2005). *Early Childhood Environment Rating Scale – Revised Edition*. New York, NY: Teachers College Press. (Updated with additional notes and a new expanded scoresheet).
- Peisner-Feinberg, E. & 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.
- Pianta, R., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS)*. Baltimore, MD: Paul H. Brookes Publishing.

- Smith, M. W., Dickinson, D. K., Sangeorge, A., & Anastasopoulos, L. (2002). *ELLCO: User's guide to the Early Language and Literacy Classroom Observation Toolkit (Research Addition)*. Baltimore, MD: Paul H. Brookes Publishing.
- Sylva, K., Siraj-Blatchford, I., & Taggart, B. (2003). *Assessing quality in the early years: Early Childhood Environment Rating Scale-Extension (ECERS-E): Four curricular subscales*. Stoke-on Trent: Trentham Books.
- Whitebook, M., Howes, C., & 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.
- Woodcock, R. W., & Johnson, M. B. (1990). *Woodcock-Johnson Psycho-Educational Battery-Revised*. Allen, TX: DLM Teaching Resources.

Early Language & Literacy Classroom Observation (ELLCO) Toolkit

I. Background Information

Author/Source

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

Publisher: Paul H. Brookes Publishing Co.
Post Office Box 10624
Baltimore, MD 21285-0624
Phone: 800-638-3775
Website: 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 pre-school 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., in 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 is intended for pre-K to third grade settings.

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 & 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 (pre-k and kindergarten version), reading instruction (school-age version), approaches to children’s writing (pre-k 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.* –Consists 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).

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Researchers, supervisors, program directors, principals, administrators, and/or teachers (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 9 hours of training is required for appropriate and responsible use.

Setting

The ELLCO is administered 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

III. Functioning of Measure

The ELLCO Toolkit was pilot tested and used in several research studies, including research conducted in over 150 pre-school classrooms for the Head Start-funded New England Quality Research Center (NEQRC) and the Literacy Environment Enrichment Project (LEEP) (Smith et al., 2008). Psychometrics reported below are based on various analyses of data from Year 4 of the NEQRC and combined data from Years 1 -3 of the LEEP.

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, interrater reliabilities of 90% and better have been consistently achieved.
- *Literacy Activities Scale:* When observers have been trained and supervised appropriately, the average interrater 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, p. 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, p. 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, p. 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, p. 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 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] & National Association for the Education of Young Children [NAEYC], 1998).

Comments

The ELLCO Pre-K and ELLCO K-3 are also available.

For more information on the ELLCO Pre-K, see the profile included in this compendium.

–The Early Language and Literacy Classroom Observation Tool, K-3 (ELLCO K-3), Research Edition, has been devised expressly for use in kindergarten through third-grade settings.” (Smith, Brady, & Clark-Chiarelli, 2008).

References and Additional Resources

- Abbott-Shim, M., & Sibley, A. (1998). *Assessment Profile of Early Childhood Programs*. Atlanta, GA: Quality Assist.
- Castro, D. C. (2005). *Early Language and Literacy Classroom Observation. Addendum for English Language Learners*. Chapel Hill. The University of North Carolina, FPG Child Development Institute.
- 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., & 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 5 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- Third Edition*. Circle Pines, MN: American Guidance Service.
- International Reading Association [IRA] & 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.
- Smith, M.W., Brady, J. P., & Anastasopoulos, L. (2008). *User’s Guide to the Early Language & Literacy Classroom Observation Pre-K Tool*. Baltimore, MD: Paul H. Brookes Publishing.
- Smith, M.W., Dickinson, D.K., Sangeorge, A., Anastasopoulos, L. (2002). *Early Language & Literacy Classroom Observation Toolkit: Research Edition*. Baltimore, MD: Paul H. Brookes Publishing.
- Snow, C.E., Burns, M.S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, D.C.: National Academy Press.

Early Language & Literacy Classroom Observation (ELLCO) Pre-K Tool

I. Background Information

Author/Source

Source: Smith, M.W., Brady, J. P., & Anastasopoulos, L. (2008). *User's Guide to the Early Language & Literacy Classroom Observation Pre-K Tool*. Baltimore, MD: Paul H. Brookes Publishing.

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

Purpose of Measure

As described by the authors:

—The ELLCO was first published in 2002 as the ELLCO Toolkit, Research Edition, and has been revised to incorporate the most recent research on early language and literacy development. Now part of a suite of products, the ELLCO Pre-K is an observation instrument that has been expressly designed for use in center-based classrooms for 3- to 5-year-old children. An additional observation instrument completes the set: The ELLCO K-3, Research Edition, is available for use in kindergarten through third grade.” (Smith, Brady, & Anastasopoulos, 2008, p. 1).

Population Measure Developed With

The authors describe the development of the ELLCO Pre-K Tool as being based on data and feedback from use of the original ELLCO Toolkit, Research Edition. They describe the development of the ELLCO Toolkit, Research Edition as follows:

—The toolkit has been pilot tested and used in several research studies since its initial development, including research conducted in more than 150 pre-school 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., in Newton, Massachusetts” (Smith, Brady, & Anastasopoulos, 2008, p. 1).

Age Range/Setting Intended For

The ELLCO Pre-K is intended for center-based settings with children ages 3- to 5-years-old.



Ways in which Measure Addresses Diversity

Item 7, *Recognizing Diversity in the Classroom*, specifically addresses diversity. The item assesses how the diversity children bring to the classroom is incorporated into ongoing curricular activities. Observers rate whether children's prior knowledge and interests are used in the classroom, efforts by teachers to create a home-school connection for all children, and whether there is an appreciation for cultural and linguistic diversity.

Key Constructs & Scoring of Measure

The ELLCO Pre-K consists of an observation instrument and a teacher interview designed to supplement the observation. The observation contains a total of 19 items, organized into five main sections.

- *Section I: Classroom Structure* (4 items)
 - Organization of the Classroom
 - Contents of the Classroom
 - Classroom Management
 - Personnel
- *Section II: Curriculum* (3 items)
 - Approaches to Curriculum
 - Opportunities for Child Choice and Initiative
 - Recognizing Diversity in the Classroom
- *Section III: The Language Environment* (4 items)
 - Discourse Climate
 - Opportunities for Extended Conversations
 - Efforts to Build Vocabulary
 - Phonological Awareness
- *Section IV: Books and Book Reading* (5 items)
 - Organization of Book Area
 - Characteristics of Books
 - Books for Learning
 - Approaches to Book Reading
 - Quality of Book Reading
- *Section V: Print and Early Writing* (3 items)
 - Early Writing Environment
 - Support for Children's Writing
 - Environmental Print

The ELLCO Pre-K is scored and tabulated so that there are two main sub-scales. Sections I and II combine to create the *General Classroom Environment* subscale. Sections III, IV, and V combine to create the *Language and Literacy* subscale. "These subscales are intentionally differentiated, with the emphasis placed on the *Language and Literacy* subscale, which contains the majority of items (12), whereas the *General Classroom Environment* subscale includes the remaining items (7)" (Smith, Brady, & Anastasopoulos, 2008, p. 1).



Major Changes to the Updated Measure

"Thanks to the widespread use of the original ELLCO Toolkit, Research Edition, and feedback from a diverse body of users, we have incorporated a range of changes. These changes serve to make the ELLCO Pre-K more focused than the original ELLCO, as well as easier to use and score. Items from the Literacy Environment Checklist and the Literacy Activities Rating Scale have been integrated into the architecture of the observation itself. The purpose of this substantial change was to make several of the observation items more robust by including the details previously gathered by the Literacy Environment Checklist and the Literacy Activities Rating Scale and to reduce some of the previous reliance on counting literacy materials and activities that tended to skew results toward classrooms that had more 'stuff,' regardless of whether or how that stuff was used" (Smith, Brady, & Anastasopoulos, 2008, p. 4).

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The authors suggest that potential users have strong background knowledge of children's language and literacy development. Additionally it is preferable for potential users to have experience teaching in pre-school classrooms. Suggested individuals for test administration are: researchers, supervisors, coaches/mentors, professional development facilitators, and teachers.

Training Required: The authors specify that training is required in order for the tool to be administered properly. Please contact Brookes Publishing at www.brookespublishing.com or 800-638-3775 for information on upcoming training opportunities.

Setting:

The ELLCO Pre-K is intended for center-based settings.

Time Needed and Cost

Time: Approximately 3 hours.

Cost: User's Guide and Toolkit: \$50.00. Please contact Brookes Publishing at www.brookespublishing.com or 800-638-3775 for information on training costs.

III. Functioning of Measure

"Given that the ELLCO Pre-K includes the content covered in the Research Edition while providing more specificity and a broader range of items, we have every reason to believe that the ELLCO Pre-K will exhibit similar, if not stronger, psychometric properties than the Research Edition. Because both instruments share the same general structure, it is appropriate to compare the ELLCO Pre-K and the Classroom



Observation from the Research Edition. The ELLCO Pre-K includes 19 items, whereas the Classroom Observation contains 14 items. If all else remains constant, the mere increase in the number of items allows for more variance, which should lead to increased reliability of the overall scale. Given the greater level of detail provided by having descriptive indicators for each scale point, we would also anticipate improved levels of interrater reliability. Clearly, this hypothesis will need to be verified empirically; therefore, we are currently collecting data on the ELLCO Pre-K in order to perform psychometric analyses, the findings of which will be provided online at <http://www.brookespublishing.com/ellco> in the near future" (Smith, Brady, & Anastasopoulos, 2008, p. 78-79).

Reliability Information

Reliability information will be available online at <http://www.brookespublishing.com/ellco> in the near future.

Validity Information

Validity information will be available online at <http://www.brookespublishing.com/ellco> in the near future.

Comments

For more information on the ELLCO Toolkit, Research Edition please refer to the profile on page 159.

References and Additional Resources

- Castro, D. C. (2005). *Early Language and Literacy Classroom Observation. Addendum for English Language Learners*. Chapel Hill. The University of North Carolina, FPG Child Development Institute.
- Smith, M. W., Brady, J. P., & Anastasopoulos, L. (2008). *Early Language & Literacy Classroom Observation Pre-K Tool*. Baltimore, MD: Brookes Publishing.
- Smith, M.W., Brady, J. P., & Anastasopoulos, L. (2008). *User's Guide to the Early Language & Literacy Classroom Observation Pre-K Tool*. Baltimore, MD: Paul H. Brookes Publishing.
- Smith, M. W., Brady, J. P., & Clark-Chiarelli, N. (2008). *User's Guide to the Early Language and Literacy Classroom Observation K-3 Tool, Research Edition*. Baltimore, MD: Brookes Publishing.
- Smith, M.W., Dickinson, D.K., Sangeorge, A., Anastasopoulos, L. (2002). *Early Language & Literacy Classroom Observation Toolkit: Research Edition*. Baltimore, MD: Paul H. Brookes Publishing.

Early Language & Literacy Classroom Observation: Addendum for English Language Learners (ELLCO: Addendum for ELL)



I. Background Information

Author/Source

Source: Castro, D. C. (2005). *Early Language and Literacy Classroom Observation. Addendum for English Language Learners*. Chapel Hill. The University of North Carolina, FPG Child Development Institute.

Publisher: Please contact Dina Castro at castro@mail.fpg.unc.edu for more information.

Purpose of Measure

As described by the authors:

"This measure has been developed as an addendum to the *Early Language and Literacy Classroom Observation Toolkit* (ELLCO), to obtain information about specific classroom practices related to promoting language and literacy development among children who are English Language Learners" (Castro, 2005, p. 2).

The *ELLCO Addendum for English Language Learners (ELLCO Addendum for ELL)* is meant to be used along with the original *ELLCO Toolkit* created by Smith et al. (2002). It is completed following the completion of each corresponding section of the *ELLCO Toolkit*. However, it can also be used as a stand-alone instrument, since it is scored separately. The measure is designed to examine classroom and instructional factors that affect the experiences of English language learners in early childhood pre-kindergarten settings. The *ELLCO Addendum for ELL* is intended for use in classrooms with children who are English language learners. Specifically the measure is targeted towards Latino dual language learners who speak Spanish at home (Castro, Espinosa, & Paéz, in press).

Population Measure Developed With

The *ELLCO Addendum for ELL* was developed in the context of the Nuestros Niños Early Language and Literacy Study. The sample consisted of 55 early childhood teachers and 193 Latino dual language learners who are part of North Carolina's More at Four (MAF) Pre-Kindergarten Program. Approximately 20% of children enrolled in the MAF program are dual language learners (Buysse, Castro, & Peisner-Feinberg, in press).

Age Range/Setting Intended For

The *ELLCO Addendum for ELL* is intended for pre-kindergarten center-based settings.



Ways in which Measure Addresses Diversity

The measure itself is intended for English Language Learner students and thus addresses diversity throughout by focusing on the unique needs of ELL children.

Key Constructs & Scoring of Measure

The *ELLCO Addendum for ELL* is meant to be used in tandem with the *ELLCO Toolkit*, but can be used as an independent measure because it is scored separately. Items in the *ELLCO Addendum for ELL* expand items in the *ELLCO Toolkit* to assess how classroom practices are addressing the particular needs of ELLs. Some new items, not included in the ELLCO Toolkit were also added. The developers suggest that "The *Addendum for English Language Learners* should be completed side by side with the ELLCO. For example, when observing the classroom to complete the Literacy Environment Checklist, have both measures at hand. First complete the information on Book Area and Book Selection for the ELLCO, and then complete the same sections for the *Addendum*" (Castro, 2005, p. 2).

- *Literacy Environment Checklist* (10 items in total)
 - Book Selection (2 items)
 - Book Use (5 items)
 - Writing Materials (1 item)
 - Writing Around the Room (2 items)
- *Classroom Observation* (8 items in total)
 - General Classroom Environment:
 - Presence and Use of Technology
 - Classroom Management Strategies
 - Presence of Books
 - Language, Literacy, and Curriculum:
 - Approaches to Book Reading
 - Approaches to Children's Writing
 - Approaches to Curriculum Integration
 - Facilitating Home Support for Language and Literacy
 - Approaches to Assessment
- *Literacy Activities Rating Scale* (8 items in total)
 - Book Reading (8 items)

The following suggestions are given for scoring the *ELLCO Addendum for ELL*. "The *Addendum* should be scored separately from the *ELLCO*, using the Addendum Score Form. Both measures use the same scoring procedures. For the Classroom Observation section, the *ELLCO* is used to establish a basis for the classroom in general, first. The *Addendum* then measures what is being done for English Language Learners *above and beyond* what the *ELLCO* measures. Therefore, the *Addendum* score will always be equal to or lower than the *ELLCO* score" (Castro, 2005, p. 2).



II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Specific prerequisites for use of the *ELLCO Addendum for ELL* are not outlined by the author. Because the *ELLCO Addendum for ELL* is done alongside the *ELLCO Toolkit* similar requirements for test administrators would be expected. 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 (Smith et al., 2002). In addition, the administrator will need to be Spanish-English bilingual. Since the instrument is still under development, suggested administrators are researchers and program evaluators.

Training Required: Those interested in using this instrument should contact the author, Dina Castro at castro@mail.fpg.unc.edu, for information about training.

Setting

The ELLCO-ELL is intended for use in pre-kindergarten to third grade center- or school-based settings serving English Language Learners.

Time Needed and Costs

Time: The *ELLCO Addendum for ELL* requires 1- 1.5 hour to be administered. When administered along with the ELLCO takes about 1-1.5 additional hours (Smith et al., 2002).

Cost: This tool is currently in its experimental phase, so cost is limited to the cost of training for instrument administration. The tool is available from the author to those who receive the training.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Data collector training entailed direct instruction, demonstration, and guided practice with feedback. Data collectors were then observed to insure that they were following proper administration procedures and had to reach the training criterion of 90% agreement with the trainer's score on classroom observations prior to data collection. Once data collection was underway, data collectors were monitored and inter-rater reliability data were gathered for 20% of the classrooms that participated in the Nuestros Niños Early Language and Literacy study. Cohen's kappa coefficients were calculated for each item on the classroom observation scale with an overall mean value of .46. Percent exact agreement was calculated for each item on the literacy environment checklist, with an overall mean value 94%. Percent exact agreement was calculated for each item on the literacy activities rating scale, with an overall mean value of 100% (Buysse, Castro & Peisner-Feinberg, in press). Internal consistency estimates are as follows: .57 for the Literacy Environment Checklist, .78



for the Classroom Observation Scale, and .30 for the Literacy Activities Rating Scale (Castro, Buysse, & Peisner-Feinberg, 2009).

Validity Information

Validity information is not available at this time.

Comments

This instrument is under development. "At this time we do not know how well this measure predicts short-term or long-term child outcomes for Spanish speaking dual language learners" (Castro et al., in press).

References and Additional Resources

Buysse, V., Castro, D. C. & Peisner-Feinberg, E. (in press). Effects of a professional development program on classroom practices and child outcomes. *Early Childhood Research Quarterly*.

Castro, D. C., Buysse, V., & Peisner-Feinberg, E. S. (2009). *Assessing Quality of Classroom Practices with the Early Language and Literacy Classroom Observation (ELLCO) Addendum for ELLs* Presented at the Biannual Meeting of the Society for Research in Child Development, Denver, CO.

Castro, D. C., Espinosa, L. M., & Paéz, M. (in press). Defining and measuring quality early childhood practices that promote dual language learners' development and learning. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Next Steps in the Measurement of Quality in Early Childhood Settings*. Baltimore, MD: Brookes Publishing.

Castro, D. C. (2005). *Early Language and Literacy Classroom Observation. Addendum for English Language Learners*. Chapel Hill. The University of North Carolina, FPG Child Development Institute.

Smith, M. W., Brady, J. P., & Anastasopoulos, L. (2008). *User's Guide to the Early Language & Literacy Classroom Observation Pre-K Tool*. Baltimore, MD: Paul H. Brookes Publishing.

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

Early Literacy Observation Tool (E-LOT)

I. Background Information

Author/Source

Source: Grehan, A.W., & Smith, L. J. (2004). *The Early Literacy Observation Tool*. Memphis, TN: The University of Memphis, Center for Research in Educational Policy.

Publisher: Education Innovations, LLC.

Purpose of Measure

As described by the authors:

The Early Literacy Observation Tool (E-LOT), a successor of the Literacy Observation Tool (LOT), is an observation instrument designed to measure research-based instructional practices, student activities, and environmental settings in early childhood classrooms where teachers are engaged in teaching the foundations of reading and other literacy processes. "The E-LOT was designed to assist schools in evaluating the effectiveness of teacher implementation of research-based teaching strategies. The E-LOT has been aligned to the National Reading Panel and National Research Council findings and captures all essential components of the Early Reading First program" (Grehan et al., 2006, p. 27).

Population Measure Developed With

The E-LOT was piloted in a rural school district of Tennessee as part of an Early Reading First evaluation. Subsequently, the E-LOT has been used in both rural and urban districts and with at-risk pre-school and kindergarten populations in school districts across the nation.

Age Range/Setting Intended For

The E-LOT is used to evaluate early childhood classrooms, particularly in pre-school and kindergarten. The E-LOT was derived from and aligns closely with the LOT (Smith, Ross, & Grehan, 2002), an instrument used to evaluate similar research-based literacy processes and practices in elementary school classrooms. The LOT has been employed and validated in multiple research and evaluation studies of Reading First and other literacy programs nationally. More information about the LOT is available from the authors.

Ways in which Measure Addresses Diversity

The instrument is designed to capture research-based instructional practices, student activities, and environmental settings regardless of context, culture, and ethnicity. Therefore, it is not biased or geared to particular cultures or subgroupings of children or teachers.

Key Constructs & Scoring of Measure

During literacy-related instruction, observers score the degree to which six components are occurring in the classroom on a 5 point scale from 0 – "Not observed" to 4 – "Extensively observed" on the *E-LOT Notes Form*. The E-LOT Notes form is completed for every 10 minutes of observed instruction (8-10 total observations are completed over the course of the observation period). The notes forms are then synthesized and summarized on the *E-LOT Data Summary Form*. The observations and a report summary are organized around the following six categories:

- *Instructional Orientation*
- *Instructional Components*
 - Concepts of Print
 - Alphabetic and Phonological Awareness
 - Fluency
 - Vocabulary and Oral Language Development
 - Development of Cognition and Text Comprehension
 - Emergent Writing
- *Assessment*
- *Learning Environment* (Scored dichotomously as either 1 "use" or 2 "non-use")
- *Visible Print Environment*
- *Materials Used*

"The subcategories of Instructional Components include the essential components of early reading identified by the National Research Council and the National Reading Panel as important in achieving effective early literacy instruction" (Grehan et al., 2006, p. 28).

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Trained observers administer the E-LOT.

Training Required: There is a three-step training process to use the E-LOT. First, observers must read a manual that describes and defines the strategies that are to be noted during the E-LOT observation. Second, the observer attends a formal training to ensure consistent observation and coding. Finally, each observer practices using the E-LOT and completes an inter-rater reliability consensus rating process with another observer of the same class to ensure consistency (Grehan et al., 2006).

Time Needed and Cost

Time: One *E-LOT Notes form* is completed during a 10-minute segment of teaching focused on early literacy development activities and learning centers. A minimum of eight E-LOT Notes forms should be completed in order to ensure reliability (i.e., at least 80 minutes of literacy instruction is observed to complete a total observation). After completing the observations, the E-LOT notes forms should be used as a

reference for completing the *E-LOT Data Summary* form. The summary process requires approximately 10 minutes for completion.

Cost: Training – The fees for training vary depending on a variety of factors including the number of people to be trained, how many tools are covered in the training (the E-LOT may be used with other standard or custom companion tools), and what role this tool may play in a larger research project being led by the Center for Research in Educational Policy (CREP) organization. Given these parameters, the standard fee is \$2,000 to train up to 40 participants + \$20 per participant for training materials including a manual + travel expenses for one trainer.

Fees paid to the site researcher – CREP personnel can conduct the actual observations using trained staff for around \$300 + travel per observation. (If CREP conducts the observations, there is no fee for training). The options most sponsors choose is for CREP to train their staff who then conduct the observations as part of their regular responsibilities or the sponsor may hire retired teachers in their own community.

Fee for processing the results – If observers record the observation data in our online system, CREP will process the results of the E-LOT for \$350 per school. This includes the cost for one annual report of results. The fee will increase by \$100 for the Center to process paper scantron sheets. Each data summary report is an easy to use document that includes tables and results of each observation for each strategy observed, but does not include any custom narrative analysis or recommendations. This additional analysis is, however, available for an additional fee.

The E-LOT, as all of our tools, is available without fee to students completing a dissertation so long as they are able to attend one of the training sessions, but the developer's agreement with the University otherwise restricts the use of the E-LOT to only those who agree to the above fee for processing the results. In addition to helping support our research center from the fees collected, this also ensures we continue to gather valuable research data from schools across the nation, all of which is included in our national norms. The University will occasionally approve a site-license fee of approximately \$100 per site where the district processes the data and creates their own report.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

In an inter-rater reliability analysis of the E-LOT administered by 30 trained observers working in pairs at 15 sites, levels of agreement (as measured by the Kappa statistic) between raters overall was between .66 - .97. The Intraclass Correlation Coefficients (ICC) revealed that one of the 83 E-LOT items, "Reviews vocabulary, including environmental print and word walls" did not have a positive ICC, indicating that there was discrepancy among raters on this item. There was low agreement on the item, "Monitors for opportunities to scaffold children's learning." The remainder

of the items had ICC values ranging from .52 – 1.00 (Huang, Franceschini, & Ross, 2007).

The E-LOT was derived from the LOT (Smith, Ross, & Grehan, 2002), which has been employed and validated in multiple research and evaluation studies of Reading First and other literacy programs nationally. A formal reliability study of LOT, conducted by Sterbinsky and Ross (2003) using Generalizability Theory, yielded mean phi coefficients across all LOT items of .75 for five observations and .82 for eight observations. Because the E-LOT overlaps extensively with LOT, including approximately 75% of the identical target categories, it is expected that E-LOT will produce comparably high reliability indices.

Validity Information

Criterion Validity

Descriptive results examining the relationship between the E-LOT and student achievement suggests a positive correlation between the scores on the observation measure and student achievement. Additionally, these descriptive results suggest that the E-LOT converges with the Classroom Observation component of the Early Language and Literacy Classroom Observation (ELLCO).

Construct Validity

The E-LOT has been aligned to the National Reading Panel and National Research Council findings and captures all essential components of the Early Reading First program” (Tennessee Reading First Formative Evaluation, 2006, p. 27).

Content Validity: The E-LOT is a successor of the LOT and the two instruments overlap extensively with approximately 75% of the target categories identical. The LOT was assessed for content validity during the developmental phase by a panel of experts, including researchers and practitioners drawn from the University of Memphis, the Memphis city schools, and the state Departments of Education in Tennessee, Louisiana, and Illinois.

References and Additional Resources

Grehan, A., Smith, L., Boyraz, G., Huang, Y., & Slawson, D. (2006) *Tennessee Reading First Formative Evaluation*. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.

Grehan, A., & Sterbinsky, A. (2005, April). *Literacy Observation Tool reliability study*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.

Grehan, A. W., & Smith, L. J. (2004). *The Early Literacy Observation Tool*. Memphis, TN: The University of Memphis, Center for Research in Educational Policy.

Huang, Y., Franceschini, L. A., & Ross, S. M. (2007). *Inter-rater reliability analysis of E-LOT*. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.

Smith, L. J., Ross, S. M. & Grehan, A. W. (2002). *Literacy Observation Tool (LOT[©])*. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.

Sterbinsky, A. & Ross, S. M. (2003). *Literacy Observation Tool Reliability Study*. Memphis, TN: Center for Research in Educational Policy, The University of Memphis.

Emlen Scales: A Packet of Scales for Measuring the Quality of Child Care From a Parent's Point of View

I. Background Information

Author/Source

Source: Emlen, A. C., Koren, P. E., & Schultze, K. H. (2000). *A packet of scales for measuring quality of child care from a parent's point of view*. Portland, OR: Regional Research Institute for Human Services, Portland State University.
<http://www.ssw.pdx.edu/focus/emlen/pgOregon.php> &
<http://www.hhs.oregonstate.edu/familypolicy/occrp/publications.html>

Purpose of Measure

As described by the authors:

The scales measure parent perceptions of the quality of their childcare arrangements. The scales are not measures of satisfaction, but provide an implicit evaluation of specific, descriptive characteristics of the care a child receives. The scales are designed to measure a parent's view of various aspects of that care, such as the warmth and interest in the child or the skill of the caregiver. The vehicle for collecting the quality data is a survey questionnaire designed to understand the work, family, and child-care context of parents' childcare decisions.

Population Measure Developed With

By the end of July 1996, the original survey had produced a composite sample of 862 parent questionnaires from more than a dozen sources inclusive of a wide range of incomes, types of jobs, and types of child care. The largest sub-sample was 264 US Bank employees who had children under age 13. Two other corporate samples were Boeing Aircraft employees using referral and counseling services and Mentor Graphics parents using an on-site child development center outstanding in quality. Members of the Association of Flight Attendants, AFL-CIO, who were living in Oregon and flying for three major airlines, provided a sample of parents with demanding work schedules, as compared to regular shifts. In addition to parents who found child care informally on their own, the study included a sample of parents who had turned to resource and referral agencies for help in finding care. Among low-income parents, samples included families receiving public childcare assistance and those who did not. All levels of household income were represented—31% with less than \$20,000, and 20% with \$75,000 or more. The amount families spent monthly on child care for all children, as a percentage of household income, provided a measure of affordability: the median spent was 9%, the middle half spent between 16% and 5%, with 29% spent by those least able to afford it.

Eight percent of the sample had a child with an emotional or behavioral problem requiring special attention. Though the overall sample consisted of current, active arrangements, many parents were facing a variety of challenges that contributed to a range in reported levels of quality of care. Two samples were selected for their recognized high quality, and, at the other extreme, were parents who had lodged complaints about care they were using.

Sixty-nine percent of the sample children were under the age of 5, with a median age of 3. Among types of child care, 89% of the parents were using paid care—38% in family day care, 35% in centers, and 8% with a grandparent. Also in the sample was care in the child's home by caregivers who were unrelated. The children were in care a median of 30 hours per week, the middle half in care between 19 and 40 hours. The middle 50% of arrangements had already lasted from 5 to 24 months—the middle 80% from 2 to 36 months.

The sample came largely from Oregon—746 (87%), 58 from Washington, 44 from California, and 14 from 8 other states. The composite sample of 862 was dispersed across 253 zip code areas.

Kyle Matchell and the Metropolitan Council on Child Care, in Kansas City, carried out a second survey in July 1997, and provided the investigators with the coded data to analyze. This sample offered an opportunity to discover if the original scales could be replicated. All parents in this sample (N = 240) found their child care through a community-based referral service—nearly three times as many in family homes as in centers, and 75% of the children were under 3 years of age (Emlen et al., 2000, p. 48). The scales resulting from that survey were strikingly similar to the original scales, and equally, or more, reliable (p. 41).

Age Range/Setting Intended For

The instrument can be used for children of all ages, in any type of child care arrangement.

Ways in which Measure Addresses Diversity

In Emlen's study, no attempt was made to classify parents in categories of racial or cultural diversity, but a few scales did touch on the fit between the diversity of children and the diversity of child care (Emlen et al., 2000, p. 39-40). See Appendix at the end of this profile.

Parent's perception of caregiver's cultural sensitivity. One measure was designed to measure the caregiver's respect for individual differences, yet in a way that could be applicable to any childcare situation or cultural difference.

Disability and special needs. Similarly, another scale measures the extent to which a child may need more attention and caregiving effort than most children. These special needs may be associated with a disability. Parents who reported, "Yes, my child has an emotional or behavioral problem that requires special attention," were 20

times more likely than other parents of children over 3 to say, "I've had caregivers who quit or let my child go because of behavioral problems" (Emlen and Wait, 1997).

Key Constructs & Scoring of Measure

There are eight scales representing conceptually and empirically distinct facets of quality of care:

- *Warmth and interest in my child* (10 items)
- *Rich activities and environment* (5 items)
- *Skilled caregiver* (8 items)
- *Talk and share information* (3 items)
- *Caregiver accepting and supportive* (4 items)
- *Child feels safe and secure* (8 items)
- *Child getting along well socially* (2 items)
- *High risk care* (11 items)

Plus a composite scale:

- *Parent scale measuring quality of child care* (15 items)

The data and method of scale construction:

The parent scales consist of evaluative statements that are simple, specific, and descriptive of the childcare experience of the parent's youngest child. Parents responded by rating how often each statement described their experience—never, rarely, sometimes, often, or always.

Based on a factor analysis of parent responses to 55 such statements, those item responses that were most highly correlated, and had a similar underlying meaning in common, were grouped together as distinguishable aspects of childcare quality from a parent's point of view. Those scales are named above.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Parents self-administer the questionnaire.

Training Required: No training required. Rated reading level: 7th grade

Setting

The instrument is appropriate for any type of care arrangement.

Time Needed and Cost

Time: Depends on total number of items in questionnaire. For quality items alone, allow 10 minutes.

Cost: To estimate cost, users should consider the following: sample size, printing, postage for mailing questionnaire and returns (unless distributed by a company or

organization), double data entry and verification, preparation of data and frequencies, data analysis, and reporting.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

There were no raters other than the parent, and no repeat measures by the same individual. Reliability was determined through analysis of internal consistency.

Internal Consistency

Factor analyses confirmed the ability of parents to discriminate levels of quality when making specific observations and judgments about their current child care" (Emlen, 2000, p. 25). This analysis differentiated distinct aspects of childcare quality and became the basis for creating a coherent set of measurement scales. The reliability of these scales was determined by the calculation of Chronbach's Alpha.

The following Cronbach's alpha coefficients were reported:

- *Warmth and interest in my child* = .93
- *Rich activities and environment* = .87
- *Skilled caregiver* = .88
- *Talk and share information* = .72
- *Caregiver accepting and supportive* = .70
- *Child feels safe and secure* = .86
- *Child getting along well socially* = .80
- *High risk care* = .73

The Kansas City replication gave additional confidence in the stability of the measures across samples.

Validity Information

Criterion Validity

Since the parent scales contained specific items that made no mention of the word "quality," it was necessary to verify that such scales were predictive of parents' judgments made explicitly about the overall "quality" of their child's care. Therefore, later in the questionnaire, parents were asked: *All things considered, how would you grade the quality of the care your child is in? (Perfect, Excellent, Good, Fair, Poor, Bad, or Awful)*. The 15-item scale's correlation to the general rating was .69. There was evidence that parents distinguish between *quality* and *satisfaction*, which tends to take circumstances into account, in addition to issues of quality. Thus, 84% said, *If I had it to do over, I would choose this care again*, but only 68% said, *The care I have is just what my child needs*.

Construct Validity

Evidence confirmed that, in expressing their idea of quality and in making their evaluative judgments, parents were not confusing or confounding quality with something else—with some other key concept, such as flexibility. In a finding

documenting construct validity, parents who rated their child care high in quality also rated their caregivers low on *caregiver flexibility*. They clearly discriminated between quality and the flexibility provided by their childcare provider.

The coherence of the findings lend support to the validity of the constructs and their measures. Those parents who were able to take advantage of the less flexible child care did so because they enjoyed high flexibility at work and within the family. The net amount of flexibility they were able to glean from any and all sources was indeed predictive of the quality of care they reported. By contrast, parents of children with an emotional or behavioral problem reported low flexibility from all sources—at work, at home, and at child care—and they reported significantly lower quality of care on all of the quality scales, as compared to other parents of children 3 years of age or older (Emlen and Weit, 1997).

Any scale measuring parent-reported quality of care was treated as the dependent variable: Emlen and colleagues (1999 and 2000) report that affordability, accessibility, flexibility, and other variables measuring child-care needs together accounted for half of the variance of quality of care measured by the 15-item scale. To assure the validity of such a prediction, an effort was made to keep the various constructs and their measures distinct. However, an accessibility measure, like *I found a caregiver who shares my values*, may be akin to quality, even though it was supposed to describe successful behavior in the childcare market.

Concurrent Validity

Two sub-samples afforded an opportunity to run a classic test of the validity of parent judgments and the quality-of-care scales based on independent criteria. The findings supported the validity of the measures. Quality-of-care scale scores were significantly higher for parents using an on-site child development center that was widely regarded as of outstanding quality—Mentor Graphic Corporation—and lower for all other center care users in the study. Also, a similar discriminating pattern was found for users of a family day care home generally acknowledged to be outstanding. In both of these comparative analyses, parents observed and correctly discriminated the level of quality of the program their child was experiencing (Emlen et al., 2000, p. 12).

Content Validity

Content validity was investigated by examining whether scores on the 15-item quality-of-care scale were statistically independent of a number of other variables: age of child, type of child care, early or late in the duration of the arrangement, the whole range in level of quality, parents' ability to read and understand the questions, and parents in all walks of life as measured by household income.

Items such as *My child feels safe and secure* can be answered for a child of any age; so, in box-plot analyses, the level and variation of reported quality did not differ significantly for infants, toddlers, or pre-schoolers through age 6. Reported quality

faded somewhat for school children, but the overall correlation was fairly low: $r = .24$.

The instrument proved applicable for any type of care for two reasons. First, similar averages and variation in quality were found in all types of care – centers, family day care, paid relative care, or paid in-home care by unrelated persons (Emlen, 1998; 1999). Also, three separate factor analyses were conducted: one for parents using center care, one for those using family day care, and one for all types of care. The factor structures were similar, producing roughly the same scales; so we were able to use the same scales for all types of care.

The scales were developed on a sample of widely distributed durations at point of interview—90% between 2 months and 3 years. The spread of quality scores was wide at all stages. Even though a sample of current arrangements will produce longer durations and higher quality on average than samples of either newly begun or terminated arrangements, many quality items were discriminating at all levels, while some items worked better at either the low end, middle, or upper level of quality.

The quality-of-care scales proved equally applicable to all income levels. Closely similar averages and variation were found at every level of household income.

The scales reported in this study covered a collection of topics that child-care professionals, parents, and the public probably would acknowledge as important aspects of child-care quality. However, no systematic test of that assumption was made. A wider pool of quality items could change the picture, but, in concept and measurement, an effort was made to differentiate quality from the other variables assumed to affect choice of care.

Comments

The research summarized in this profile was funded by a grant from the Child Care Bureau, Administration for Children and Families, Department of Health and Human Services, with support also from Portland State University and other participating institutions within the Oregon Child Care Research Partnership.

References and Additional Resources

Emlen, A. C., Koren, P. E., & Schultze, K. H. (2000). *A packet of scales for measuring quality of child care from a parent's point of view*. Portland, OR: Regional Research Institute for Human Services, Portland State University.
<http://www.ssw.pdx.edu/focus/emlen/pgOregon.php> &
<http://www.hhs.oregonstate.edu/familypolicy/occrp/publications.html>

Emlen, A. C., Koren, P. E. & Schultze, K. H. (1999). *From a Parent's Point of View: Measuring the Quality of Child Care: A Final Report*. Portland, OR: Regional Research Institute for Human Services, Portland State University.
<http://www.ssw.pdx.edu/focus/emlen/pgOregon.php> &
<http://www.hhs.oregonstate.edu/familypolicy/occrp/publications.html>

Emlen, A. C. (1998). "From a Parent's Point of View: Flexibility, Income, and Quality of Care" Background paper for SEED Consortium Conference: Child Care in the New Policy Context. Session: New Perspectives on Child Care Quality. NIH Campus, Bethesda, MD, April 30 and May 1, 1999.
<http://www.ssw.pdx.edu/focus/emlen/pgOregon.php/pdfBethesda1988.pdf>

Emlen, A. C. & Weit, K. (1997). "Quality of Care for Children with a Disability," 1997 Conference Proceedings, Building on Family Strengths: Research and Services in Support of Children and Families. Portland, OR: Research and Training Center on Family Support and Children's Mental Health, Portland State University.

APPENDIX

The Emlen Scales Measuring the Quality of Child Care From a Parent's Point of View

Scales Addressing Issues of Diversity

Parent perception of caregiver's cultural sensitivity (8 items)

Alpha=.88

My child is treated with respect.
The caregiver makes an effort to get to know my child.
The caregiver accepts my child for who she (he) is.
The caregiver takes an interest in my child.
My child feels accepted by the caregiver.
I feel welcomed by the caregiver.
My caregiver accepts the way I raise my child.
My caregiver is supportive of me as a parent.

Child's special needs (5 items)

Alpha=.78

My child needs more attention than most children.
My child's special needs require a lot of extra effort.
My caregiver feels that my child's needs are quite demanding.
I've had caregivers who quit or let my child go because of behavioral problems.
My child can be quite difficult to handle.

These special needs may be related to a disability:

My child has a physical disability that requires special attention.
My child has a health need that requires extra attention.
My child has an emotional or behavioral problem that requires special attention.
My child has a learning disability that requires specialized approaches.

Combining the above into one scale created a 9-item scale, Alpha=.75

Continuity of care (4 items)

Alpha=.79

My child has been in a familiar place with people he (she) knows.
My child has had stability in her/his child-care relationships.

There has been too much turnover in my child's caregivers. (-)
How many months has your child been in this arrangement?²

Scales Measuring Aspects of Child-Care Quality

(From the Kansas City Replication (N=240) (Emlen et al., 2000)

Caregiver's warmth and interest in my child (6 items) Alpha=.92

My caregiver is happy to see my child.
The caregiver is warm and affectionate toward my child.
My child is treated with respect.
The caregiver takes an interest in my child.
My child gets a lot of individual attention.
The caregiver seems happy and content.

Rich environment and activities (5 items) Alpha=.91

There are lots of creative activities going on.
It's an interesting place for my child.
There are plenty of toys, books, pictures, and music for my child.
In care, my child has many natural learning experiences.
The caregiver provides activities that are just right for my child.

Caregiver's skill (3 items) Alpha=.80

The caregiver changes activities in response to my child's needs.
My caregiver knows a lot about children and their needs.
My caregiver is open to new information and learning.

Supportive parent-caregiver relationship (6 items) Alpha=.84

My caregiver and I share information.
We've talked about how to deal with problems that might arise.
My caregiver is supportive of me as a parent.
My caregiver accepts the way I want to raise my child.
I'm free to drop in whenever I wish.
I feel welcomed by the caregiver.

Child feels happy, safe, and secure (6 items) Alpha=.85

My child feels safe and secure.
My child has been happy in this arrangement.
My child has been irritable since being in this arrangement. (-)
My child feels accepted by the caregiver.
My child likes the caregiver.
My child feels isolated and alone in care. (-)

Risks to health, safety, and well-being (10 items) Alpha=.85

My child is safe with this caregiver. (-)
There are too many children being cared for at the same time.

² This duration item was added in the Kansas City replication (Emlen et al., 2000, p.34)

The caregiver needs more help with the children.
The caregiver gets impatient with my child.
The children seem out of control.
The conditions are unsanitary.
The children watch too much TV.
It's a healthy place for my child. (-)
I worry about bad things happening to my child in care.
Dangerous things are kept out of reach. (-)

Child getting along well socially (from the original sample, N=862) (2 items) Alpha=.80

My child gets along well with the other children in care.
My child likes the other children.

15-Item Parent Scale Measuring Quality of Child Care, N=862. Alpha=.91

My child feels safe and secure in care.
The caregiver is warm and affectionate toward my child.
It's a healthy place for my child.
My child is treated with respect.
My child is safe with this caregiver.
My child gets a lot of individual attention.
My caregiver and I share information.
My caregiver is open to new information and learning.
My caregiver shows she (he) knows a lot about children and their needs.
The caregiver handles discipline matters easily without being harsh.
My child likes the caregiver.
My caregiver is supportive of me as a parent.
There are a lot of creative activities going on.
It's an interesting place for my child.
My caregiver is happy to see my child.

Scales Measuring Sources of Parent's Flexibility

Work Flexibility (5 items) Alpha=.74

Our work schedule keeps changing. (-)
My shift and work schedule cause extra stress for me and my child. (-)
Where I work it's difficult to deal with child care problems during working hours. (-)
My life is hectic. (-)
I find it difficult to balance work and family. (-)

Family Flexibility (4 items) Alpha=.78

I have someone I can share home and care responsibilities with.
I'm on my own in raising my child. (-)
Do you have a spouse or partner who is employed?
1. No spouse or partner. 2. Spouse or partner employed full time.
3. Spouse or partner employed part time. 4. Spouse not employed.
In your family, who takes responsibility for child-care arrangements?
1. I do completely. 2. Mostly I do. 3. Equally shared with spouse or other.

4. *Mostly spouse or other does.*

5. *Spouse or other does completely.*

Caregiver Flexibility (4 items)

Alpha=.81

My caregiver understands my job and what goes on for me at work.

My caregiver is willing to work with me about my schedule.

I rely on my caregiver to be flexible about my hours.

I can count on my caregiver when I can't be there.

Scales Measuring Accessibility of Child Care, Options, and Choice

Found a caregiver who shares my values (3 items)

Alpha=.80

I found a caregiver who shares my values.

I like the way my caregiver views the world.

My caregiver and I see eye to eye on most things.

Child-care options in the neighborhood (5 items)

Alpha=.77

I've had difficulty finding the child care I want. (-)

There are good choices for child care where I live.

In my neighborhood child care is hard to find. (-)

When I made this arrangement, I had more than one option.

In choosing child care, I've felt I had to take whatever I could get. (-)

Transportation a problem (4 items)

Alpha=.61

My child care is too far from home. (-)

Transportation is a big problem for me. (-)

Getting to work is a long commute. (-)

Getting my child places is difficult for me. (-)

Scales Measuring Perceived Affordability

Difficulty paying for child care (3 items)

Alpha=.78

I have difficulty paying for child care.

I worry about making ends meet.

The cost of child care prevents me from getting the kind I want.

Have some choice about how much to work (2 items)

Alpha=.84

I have some choice about whether to work or how much.

I can (or could) afford to work part time.

Environment and Policy Assessment and Observation (EPAO)

I. Background Information

Author/Source

Source: Ball, S. C., Benjamin, S. E., Hales, D. P., Marks, J., McWilliams, C. P., & Ward, D. S. (2005). *The Environment and Policy Assessment and Observation (EPAO) child care nutrition and physical activity instrument*. Center for Health Promotion and Disease Prevention, University of North Carolina at Chapel Hill. Revised summer 2006.

Ward, D. S., Hales, D., Haverly, K., Marks, J., Benjamin, S. E., Ball, S. C., & Trost, S. (2008). An instrument to assess the besogenic environment of child care centers. *American Journal of Health Behavior*, 32, 380-386.

Publisher: This measure is currently unpublished.

Purpose of Measure

As described by the authors:

"The purpose of the EPAO is to objectively and effectively describe the nutrition and physical activity environment and practices of child care facilities" (Ball et al., 2005, p. 2).

"The EPAO instrument was created to evaluate the Nutrition and Physical Activity Self-Assessment for Child Care (NAP SACC) program, an environmental nutrition and physical activity intervention in child care. . . .The NAP SACC program includes, as the centerpiece of the intervention, a self-assessment component. Content areas for the self-assessment were created following an extensive review of the nutrition and physical activity literature, recommendations and standards from credible organizations, as well as input from a number of experts in the field of measurement, nutrition and physical activity, and child care. Because the self-assessment was designed as part of the intervention, an outcome measure was created that could measure the impact of the intervention on the centers' environments. The EPAO is an expansion of the self-assessment into a tool that is executed by objective, trained field observers through direct observation and document review" (Ward et al., 2008, p. 381).

Population Measure Developed With

The EPAO was pilot tested in a child care center in North Carolina by three trained observers over the period of one day. Following the preliminary field observation the EPAO was completed in 9 North Carolina child care centers between fall 2005 and

spring 2006. These observations were the basis for the data on inter-observer agreement.

Age Range/Setting Intended For

The EPAO is designed to be used for pre-school aged children, center staff, and center director in a child care setting.

Ways in which Measure Addresses Diversity

In the Menu Review-Weekly Menus section there is a question asking whether weekly menus at the center include foods from a variety of cultures.

Key Constructs & Scoring of Measure

"The EPAO protocol consists of one full-day visit to a child care center and includes direct observation and document review activities. Observations consist of food and beverages served, staff-child meal interactions, active play time opportunities (indoor and outdoor) and sedentary behavior opportunities, staff support for nutrition and physical activity, the physical activity environment (e.g., fixed and portable equipment and outdoor space) and the nutrition environment (e.g., how children are fed). The document review involves an evaluation of the teacher's lesson plan for that week, past or future fundraising documents, menus for up to one month that include the week of the visit, parent handbook, staff handbook, most recent playground safety check, physical activity and/or nutrition training documents, physical activity and/or nutrition curricula, and written nutrition and physical activity policies (e.g. food brought from home, expectations for going outside, and education/training for staff)" (Ward et al., 2008, p. 381).

NAP SACC EPAO Observation

Recording Form 1: This portion of the observation is broken into seven sections with 64 items.

- *Eating occasions- Foods* (15 items): Observers note when meals and snacks are served and the number of times throughout the day that healthy and unhealthy foods and beverages are served.
- *Eating Occasions- Beverages* (7 items): Observers assess the availability of water, milk, and sugary drinks throughout the day
- *Eating Occasions – Staff Behaviors* (11 items): Observers record staff behaviors when serving food, eating food with the children, or eating food separate from but in front of the children. Observers also note whether any nutrition education is provided by staff.
- *Physical Activity – Child Behaviors* (5 items): Observers assess the amount of time and type of physical activity in which children participate throughout the day in addition to availability of water during physical activity.
- *Sedentary Activities - Child* (7 items): Items include number of times children are observed seated for more than 30 minutes, whether there is a television, computer, or video games available, and how often children are engaging in TV viewing, and game playing.

- *Physical Activity – Staff Behaviors* (7 items): Observers assess whether staff use restriction or increase of active play time as punishment or reward. In addition, observers note staff participation with children in physical activities and staff attitudes towards physical activity in the presence of children.
- *Center Environment* (12 items): Observers record features of the general environment such as play structures (both indoor and outdoor) available to children.

NAP SACC EPAO Document Review Form

The recording form is divided into 4 sections

- *Section 1: Menu Review – Observed Foods & Beverages* (14 items)
This section assesses consistency between the posted food menu and actual foods served.
- *Section 2: Menu Review – Weekly Menus* (2 items)
This section assesses whether weekly menus include foods from a variety of cultures.
- *Section 3: Guideline Reviews* (5 items)
This section assesses foods that are offered outside of regular meals and snacks, nutrition policy, play environment, and the centers' physical activity environment.
- *Section 4: Training & Curriculum Review* (8 items)
This section assesses staff training in and offering of nutrition and physical activity to parents and children.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Trained observers administer the measure.

Training Required: Observers for the NAP SACC study are trained during a one-day intensive workshop held by the developers of the EPAO instrument. Included in this training are lessons on general observation techniques, types of play equipment and space, instruction and demonstration of record keeping and form completion, and an overview of general child care center rules, regulations, and state mandates. A mock observation in a child care center is completed by all trainees for practice. For observers to be certified, they have to obtain 85% agreement with a gold standard observer. Periodic retraining occurs for reliability purposes (Ward et al., 2008).

In the protocol and procedures document for the EPAO, it states that data collectors undergo a one-day extensive training. Training includes: 1) General observation techniques, 2) A review of the EPAO instrument and its uses, 3) A lesson on menu review, 4) Lessons on interview techniques and procedures, 5) Instruction and demonstration of record keeping and form completion, and 6) A mock-observation to be completed alongside a gold standard observer (Ball et al., 2005).

Setting

The EPAO is administered in child care centers.

Time Needed and Cost

Time: –The observation is conducted from early in the morning before the first eating occasion of the day and continues until all of the children in that classroom leave for the day or until multiple classrooms collapse together at the end of the day. . .

The document review is completed subsequent to the observation, preferably on the same day, and includes an evaluation of the written documents described above. The document review is completed to gather additional information beyond that possible with a one day observation” (Ward et al., 2008, p. 381).

Cost: A copy of the observation and document review items are available from the primary author by request. Diane Ward, Professor, The University of North Carolina, Schools of Public Health and Medicine, Department of Nutrition, Chapel Hill, NC. dsward@email.unc.edu.

III. Functioning of Measure

Reliability Evidence

Inter-rater Reliability

Mean agreement between observer pairs: (Ward et al., 2008)
87.28% (SD = 4.21) for the observation component
79.29% (SD = 7.43) for the document review component

More information will be available in the near future.

Validity Evidence

Concurrent Validity

The eight physical activity environment subscales and the EPAO PA environment total score are significantly related to physical activity of children during child care (Bower et al., 2008).

Discriminant Validity

Significant differences in physical activity behavior are evident between centers with high and low EPAO scores (Bower et al., 2008).

Predictive Validity

Natural changes in physical activity are moderate to strongly correlated with changes in EPAO physical activity environment scores (Hales & Ward, 2008). Centers where the environment got better (increase in EPAO score) over time showed moderate to large increases in moderate and vigorous physical activity (Hales & Ward, 2008).

Content Validity

"Content areas for the self-assessment were created following an extensive review of the nutrition and physical activity literature, recommendations and standards from credible organizations, as well as input from a number of experts in the field of measurement, nutrition and physical activity, and child care" (Ward et al., 2008, p. 381).

References and Additional Resources

- Ball, S. C., Benjamin, S. E., Hales, D. P., Marks, J., McWilliams, C. P., & Ward, D. S. (2005). *The Environment and Policy Assessment and Observation (EPAO) child care nutrition and physical activity instrument*. Center for Health Promotion and Disease Prevention, University of North Carolina at Chapel Hill. Revised summer 2006.
- Bower J. K., Hales, D. P., Tate, D. F., Rubin, D. A., Benjamin, S. E., & Ward, D. S. (2008). Relationships between the child care environment and physical activity behavior in children. *American Journal of Preventive Medicine, 34*, 23-29.
- Hales, D. P., & Ward, D. S. (2008). *The impact of environmental changes in childcare on children's physical activity*. Poster presented at the International Congress on Physical Activity and Public Health, Amsterdam, the Netherlands.
- Ward, D. S., Hales, D., Haverly, K., Marks, J., Benjamin, S. E., Ball, S. C., & Trost, S. (2008). An instrument to assess the besogenic environment of child care centers. *American Journal of Health Behavior, 32*, 380-386.

Family Child Care Environment Rating Scale-Revised Edition (FCCERS-R)

I. Background Information

Author/Source

Source: Harms, T., Cryer, D., & Clifford, R. M. (2007). *Family Child Care Environment Rating Scale – Revised Edition*. New York, NY: Teachers College Press.

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

Purpose of Measure

As described by the authors:

The Family Child Care Environment Rating Scale – Revised Edition (FCCERS-R) measures global quality of care that is provided in an individual's home for a small group of children. The instrument may be used for providers' self-evaluation, monitoring by licensing or other agency staff, training, technical assistance, and for research and evaluation.

The FCCERS is a revision of the Family Day Care Rating Scale (FDCRS; Harms & Clifford, 1989). It contains items "to assess provisions in the environment for a wide age range, and to ensure protection of children's health and safety, appropriate stimulation through language and activities, and warm, supportive interaction" (Harms, Cryer, & Clifford, 2007, p.1). The scale focuses on the conditions that the authors have identified as being important for promoting children's positive outcomes.

A Spanish language edition of the FCCERS-R is also available.

Population Measure Developed With

The original version of the Scale was developed in consultation with family child care providers in North Carolina and subsequent observation of providers in North Carolina and other states. The current version was based on extensive experience in training observers in many states in the US and in consultation with providers and trainers with experience in settings serving children from diverse racial, ethnic, income and special needs perspectives.

Age Range/Setting Intended For

The FCCERS-R was developed for use in home-based child care settings serving children from birth through elementary school.



Ways in which Measure Addresses Diversity

- Indoor space used for child care (item # 1) assesses whether the space is accessible to children and adults with disabilities.
- Arrangement of indoor space for child care (item #4) assesses accessibility of spaces for children with disabilities.
- Using books (item #15) assesses whether there is a wide selection of books accessible that reflect different races, ages and abilities.
- Music and movement (item #18) assesses whether music from different cultures and in different languages is used with the children.
- Dramatic play (item # 20) assesses whether props such as dolls and dress up clothes are provided to represent diversity of cultures and abilities.
- Promoting acceptance of diversity (item # 24) assesses whether the materials and activities represent and portray positively different races, cultures, ages, gender and abilities.
- Active physical play (item #26) assesses whether materials and equipment for active play are suitable for children with disabilities.
- Provisions for children with disabilities (item #34) 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 the family child care provider interacts with parents and specialists to plan for meeting the child's needs.

Key Constructs & Scoring of Measure

The scale consists of 38 items categorized into seven subscales. Items are scored on a 7-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. It should be noted that indicators under inadequate are scored in the opposite direction from indicators at the higher levels.

- *Space and Furnishings* (6 items)
 - Indoor space used for child care
 - Furnishings for routine care, play, and learning
 - Provision for relaxation and comfort
 - Arrangement of indoor space for child care
 - Display for children
 - Space for privacy
- *Personal Care Routines* (6 items)
 - Greeting/departing
 - Nap/rest
 - Meals/snacks
 - Diapering/toileting
 - Health practices



- Safety practices
- *Listening and Talking* (3 items)
 - Helping children understand
 - Helping children use language
 - Using books
- *Activities* (11 items)
 - Fine motor
 - Art
 - Music and movement
 - Blocks
 - Dramatic play
 - Math/number
 - Nature/science
 - Sand and water play
 - Promoting acceptance of diversity
 - Use of T.V., video, and/or computer
 - Active physical play
- *Interaction* (4 items)
 - Supervision of play and learning
 - Provider-child interaction
 - Discipline
 - Interactions among children
- *Program Structure* (4 items)
 - Schedule
 - Free play
 - Group time
 - Provisions for children with disabilities
- *Parents and Provider* (4 items)
 - Provisions for parents
 - Balancing personal and caregiving responsibilities
 - Opportunities for professional growth
 - Provisions for professional needs

Comments

The FCCERS-R contains Notes for Clarification for many of the items that define the terms used in the item and clarify specific scoring requirements for the indicators that comprise the item.

The FCCERS-R and other ERS instruments are also available in electronic form for use on Tablet PC machines through software package developed by the Branagh Information Group (<http://www.ersdata.com>) under license from Teachers College Press. This package is most appropriate for medium and large scale users.



II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The FCCERS-R should be used by a trained observer at a time when children are awake and active. The authors recommend that 3 hours or more be spent observing in the home. An additional 20 – 30 minutes is needed to ask the provider questions to help score indicators that were not observed. The authors provide specific instructions for administering the scale and for conducting the observation in a way that minimizes the impact of the observer on the family child care home. The observer should have the FCCERS-R book with her/him while in the home. Care is urged by the authors to avoid conflicts of interest by observers as this has been shown to affect reliability and accuracy of scores on similar instruments.

Training Required: The authors recommend that observers participate in a training sequence led by an experienced FCCERS-R trainer. 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).

Setting

Observations are made in the family child care home.

Time Needed and Cost

Time: The FCCERS-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 & Riley, 2003, p. xiv). The authors recommend that at least 3 hours be spent observing in the classroom and note that spending more than 3 hours observing is preferable. An additional 20 – 30 minutes is needed to ask the teacher questions to help score indicators that were not observed. "A valid observation requires the presence of a representative sample of children enrolled" including children from each age group enrolled (Harms, Cryer & Clifford, 2007, p. 7).

Cost: All materials are available through Teachers College Press.

Manuals

FCCERS-R, with additional notes and expanded score sheet 2007 \$19.95

Spanish FCCERS-R 2009 \$19.95

Video Training Packages

2007, VHS \$59.00

2007, DVD \$59.00



III. Functioning of Measure

Reliability Information from Manual

Inter-rater Reliability

Inter-rater reliability was established in a field test with 8 data collectors who conducted paired observations in 45 family child care homes in North Carolina. The sample included a range of settings with different quality ratings (as rated by the North Carolina Star Rating System) and serving children of varying ages.

The scale has 460 indicators on 38 items that are grouped into 7 subscales and combined for a total score. Reliability at each of these levels is described below:

- Indicator reliability – the mean percent agreement across indicators was 88.5%
- Item reliability – the average item reliability, in which two observers scored an item within one point of each other, was 88.44%. Twelve items had weighted kappas under .60.
- Subscale and total scale reliability – The percent agreement within one point ranged between 80% and just over 90% for the seven subscales and was 88.44% for the total scale. The weighted kappas ranged from .62 to .77 for the seven subscales and was .71 for the total scale.

Internal Consistency

Chronbach's alpha was used to examine the internal consistency of the subscales and total scale. The Alphas are presented below:

<i>Space and Furnishings</i>	.71
<i>Personal Care Routines</i>	.46
<i>Listening and Talking</i>	.83
<i>Activities</i>	.88
<i>Interaction</i>	.84
<i>Program Structure</i>	.62
<i>Parents and Provider</i>	.39
<i>Total</i>	.90

Because of the high alpha for the Total Scale Score, the authors note that the FCCERS-R "appears to be a measure of global quality that reflects a single major construct" (Harms, Cryer & Clifford, 2007, p. 5). The authors recommend that the subscale scores not be used in research, though they are "quite useful both for practitioners and for those providing technical assistance in the field" (Harms, Cryer & Clifford, 2007, p. 5).



Test-Retest Reliability

The FCCERS-R was conducted twice in 20 sites. At time 1, the overall mean score was 3.32; at time 2, the mean score was 3.39. At the item level, the retest agreement within one point was 80.8%. The correlation between time 1 and time 2 observations was .73.

Validity Information from Manual

Predictive Validity

The authors do not provide direct evidence of the predictive validity of the FCCERS-R. However, they note that the scale is part of a series of environmental rating scales and that "environmental quality as defined by these instruments is predictive of child outcomes both during the period of time children are in these environments and as they move into school" (Harms, Cryer & Clifford, 2007, p. 2). The authors provide a sampling of references demonstrating predictive validity (Burchinal, Roberts, Nabors, & Bryant, 1996; Burchinal et al., 2000; Helburn, 1995; Peisner-Feinberg et al., 2001).

References and Additional Resources

- Burchinal, M., Roberts, J., Nabors, L., & Bryant, D. (1996). Quality of center child care and infant cognitive and language development. *Child Development, 67*, 606-620.
- Burchinal, M., Roberts, J., Riggins, R., Jr., Ziesel, S. A., Needbe, E., & Bryant, D. (2000). Relating quality of center-based child care to early cognitive and language development longitudinally. *Child Development, 71*, 339-357.
- Cryer, D., Harms, T., & Riley, C. (2003). *Early Childhood Environment Rating Scale*. New York: Teacher College Press.
- Harms, T. & Clifford, R.M. (1989). *Family Day Care Rating Scale*. New York, NY: Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R. M. (2007). *Family Child Care Environment Rating Scale – Revised Edition*. New York, NY: Teachers College Press.
- Helburn, S. (Ed.). (1995). *Cost, Quality, and Child Outcomes in Child Care Centers: Technical Report*. Denver: University of Colorado, Department of Economics, Center for Research in Economic Social Policy.
- Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Yazejian, N. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development, 72*, 1534-1553.

The Individualized Classroom Assessment Scoring System (inCLASS)

I. Background Information

Author/Source

Source: Downer, J. T., Booren, L. M., Lima, O. A., Luckner, A. E., & Pianta, R. C. (2009). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. *Early Childhood Research Quarterly*.

Publisher: This measure is currently unpublished.

Purpose of Measure

The Individualized Classroom Assessment Scoring System (inCLASS) is an observational assessment of children's competent interactions with adults, peers, and learning activities in pre-school and kindergarten classrooms. The purpose of this measure is to examine children's development and readiness to learn using a psychometrically sound measurement tool that is context-specific within a naturalistic classroom environment (Downer, Booren, Lima, Luckner, & Pianta, 2009a). As such, the inCLASS is an observational system that examines children's behavior across all classroom settings, having the potential to inform and evaluate classroom interventions. The aim is for both researchers and teachers to use the inCLASS in order to better understand children's adjustment to key interactions within early classroom environments.

The inCLASS measures the quality of classroom interactions at the level of an *individual child* and was designed to complement the Classroom Assessment Scoring System (CLASS; Pianta, La Paro, & Hamre, 2008), which measures the quality of interactions as the *average experience of all children* in the classroom.

Population Measure Developed With

The inCLASS was originally tested in a pilot study conducted by the University of Virginia with a sample of 164 children ages 3 to 5 years old ($M = 49$ months) across 44 pre-school classrooms. Four children from each classroom were randomly selected from those who consented. There were 90 girls and 74 boys in the sample. Children were primarily Caucasian (89% at baseline) and upper-middle-class (49% of parents reported an annual income of \$85,001 or more). There were 39 lead teachers; 95% of teachers were Caucasian and 23% had a Bachelor's degree. The average class size was 15.36 children. Classrooms were visited twice in the fall of 2006, with visits typically one week apart, followed by another two site visits in the spring of 2007. During each visit, children were observed in alternating 15-minute cycles

across the full morning, for an average of four cycles per visit and 16 cycles across the year. Twenty percent of the observations were coded by two trained observers for reliability purposes (mean agreement = 88.5% within 1; Downer, Booren, Lima, Luckner, & Pianta, 2009b). Classroom teachers also completed a packet of rating scales for each participating child in both the fall and spring, and parents completed a family demographics survey.

The inCLASS also recently completed a field study in Los Angeles to further validate the instrument with a more diverse sample and to follow children through kindergarten entry. The initial baseline sample consisted of approximately 341 children (170 girls and 171 boys, primarily Hispanic/Latino) and 84 lead teachers (2 males) across 100 pre-school classrooms (35 classrooms were Head Start). The inCLASS is also being utilized currently in several other research projects to further validate the measure. One study in Miami focuses on developing a new early childhood science measure. Another study in New Jersey is an intervention based on professional development and classroom consultation for teachers, which focuses on how to resolve early problem behaviors. Lastly, the inCLASS is being used as an evaluation tool for a 9-state randomized, controlled evaluation of a professional development program to promote language and early literacy (Downer et al., 2009b).

Age Range/Setting Intended For

The inCLASS is designed to be used in pre-school and kindergarten classrooms across all settings with children between the ages of 3 and 5 years.

Ways in which Measure Addresses Diversity

The instrument is being tested with a diverse sample of young children in multiple areas of the country, as noted above.

Key Constructs & Scoring of Measure

The inCLASS currently addresses 10 dimensions of children's behavior within three developmental domains:

- *Teacher Interactions*
 - Positive Engagement
 - Teacher Conflict
 - Teacher Communication
- *Peer Interactions*
 - Peer Sociability
 - Peer Conflict
 - Peer Assertiveness
 - Peer Communication
- *Task Orientation*
 - Engagement
 - Self-Reliance
 - Behavior Control

Each dimension is scored individually on a 7-point scale based on the observed quality and duration of specific behavioral markers described in the user's manual (e.g., noncompliance, leadership, etc.). Scores are categorized as "Low" (1, 2), "Mid" (3, 4, 5), and "High" (6, 7). The inCLASS is not a checklist. Rather, the inCLASS dimensions should be viewed as holistic descriptions of children that fall in the low, mid, and high range. Along with these 10 dimensions, information on the classroom activity setting is also collected as part of the inCLASS measure (e.g., whole group, free choice/centers, etc.).

II. Administration of Measure

Who Administers Measure/Training Required

The inCLASS was designed for researchers who want a child-focused classroom observation tool that is grounded in empirical research on important prosocial and task oriented child interactions. The inCLASS is also being tested with teachers who want to learn more about their students' behaviors in the classroom by using an authentic observational assessment.

The inCLASS is still in early development and is not currently being offered on a wide-spread basis. Extensive training is required to become a reliable inCLASS observer, which includes a 2-day training session of coding and discussing video training clips. To be reliable, observers must code within one of the mastercode on at least 80% of the dimensions. Currently, the inCLASS trainings can only be conducted by University of Virginia faculty who are certified as official trainers. An online training protocol is being developed for teachers.

Setting

The inCLASS was designed to be used across all settings in pre-school and kindergarten classrooms.

Time Needed and Cost

Time: A minimum of one morning is required to complete the inCLASS for three to four children in a classroom by alternating cycles. The instrument is administered across a series of 15-minute observation cycles, with 10 minutes to observe and 5 minutes to score. Observers typically select 4 children in the classroom, observe and score one child per 15-minute cycle, and then rotate to the next child for the subsequent cycle. Across a full 4-hour morning, an observer can typically complete 4 cycles per child for a total of 16 observation cycles.

Cost: Not yet available.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Inter-rater reliability is investigated in two ways: initial inCLASS training and double coded live observations. At the end of inCLASS training, all observers are required to watch five reliability clips for which they had to be within one from the mastercode on 80% of the dimensions to be reliable. All trainees are required to be reliable before going out into the field. During the pilot study, all coders were within one of the mastercode for 85% of the dimensions across all five training videos (a range of 74 to 92% across the 9 dimensions), with a good intraclass correlation of .65 (Downer et al., 2009b). Similar reliability was observed for training during the field study.

Inter-rater reliability for live observations was calculated across 20% of all live classroom observations where two coders observed and rated the same children. During the pilot study, observers were within one of each others' codes 87% of the time in the fall and 90% in the spring (a range of 71 to 99% across the 9 dimensions). An intraclass correlation was also calculated across all dimensions and reached .84 in the fall and .85 in the spring, both within the excellent range according to standards in the field. Intraclass correlations at the domain and dimension levels ranged from moderate to excellent (.46-.84; Downer et al., 2009b).

Validity Information

Face Validity

The inCLASS was developed based on an extensive literature review of the important cognitive and socioemotional skills developing during the pre-school period, which predict children's later social and academic performance in school. The choice of dimensions was additionally informed by a review of constructs assessed in other observational, teacher report, and direct assessment instruments currently used in child care and research. Finally, the operational definitions of the dimensions were specified through extensive piloting. Consultation with early childhood practitioners, as well as researchers with expertise in child development and school readiness, confirmed that the inCLASS measures aspects of children's classroom behavior that impact their school performance and socioemotional competency, suggesting considerable face validity (Downer et al., 2009b).

Criterion-Related Validity

To establish criterion-related validity from the pilot study, bivariate correlations were conducted to compare inCLASS Teacher Interactions, Peer Interactions, Task Orientation, and Conflict Interactions factor scores with teacher ratings from several established measures. Results from the correlational analyses generally supported the concurrent and discriminant validity of the inCLASS domains (Downer et al., 2009a).

Within the Teacher Interactions domain, observations of the target child's interactions with the teacher were positively correlated with teacher ratings of closeness with that child ($r = 0.25, p \leq .01$) and teacher reports of assertiveness ($r = 0.23, p \leq .01$).

Within the Peer Interactions domain, significant correlations were found between inCLASS observations and teacher ratings of assertiveness ($r = 0.41, p \leq .01$), social communication ($r = 0.23, p \leq .01$), language and literacy skills ($r = 0.31, p \leq .01$), and social skills ($r = 0.16, p \leq .05$). Interestingly, observations of Peer Interactions were also positively related to teacher ratings of conflict ($r = 0.19, p \leq .05$), and negatively correlated with teacher ratings of frustration tolerance ($r = -0.24, p \leq .01$).

Within the Task Orientation domain, children who were observed as having higher quality interactions with classroom tasks and activities were also rated more highly ($p \leq .01$) by their teacher on a host of skills, including task orientation ($r = 0.26$) and language and literacy skills ($r = 0.30$). In addition, ratings on the Task Orientation domain were negatively related to teacher ratings of problem behavior ($r = -0.28$). Observations within the Conflict Interactions domain were also significantly associated with similar teacher ratings. Moderately positive correlations were observed for reports of conflict and problem behaviors (r 's = 0.53, 0.41, p 's $\leq .001$), whereas other significant associations in this domain were negative, such as frustration tolerance ($r = -0.50$). Unexpectedly, Conflict Interactions were positively associated with teacher ratings of assertiveness ($r = .17, p \leq .05$; Downer et al., 2009a).

Similar concurrent and discriminant validity findings have been found using inCLASS field study data (Downer, Booren, Luckner, & Pianta, 2009).

Predictive Validity

Using the pilot data, the association between the inCLASS and children's outcomes was assessed after adjusting for a variety of covariates, including child gender, age and maternal education. Due to the nested nature of these data, hierarchical linear models were conducted using fall observations to predict spring teacher ratings. Unconditional models were run first to establish that there was significant variance at the teacher level (ICC's ranged from 0.10 - 0.34). Next, conditional hierarchical linear models were run while controlling for children's gender, age and maternal education. As was observed in the comparison with concurrent ratings, observed Teacher interactions in the Fall significantly predicted Spring teacher ratings of closeness and assertiveness. The Peer observations also significantly predicted peer-relevant ratings by the teacher, such as social skills and assertiveness. The Task observations significantly predicted teacher ratings of task orientation, language/literacy skills, and emotional regulation. Some cross-domain relationships were also observed: for example, Task observations predicted teacher-rated closeness, assertiveness and social skills. These findings suggest that observations of children's classroom interactions are related to teacher ratings of similar constructs (Booren, Abry, Luckner, Yoder, Lima, Downer, & Pianta, 2008; Downer et al., 2009b).

Preliminary predictive validity was also observed in the inCLASS field study. Using two-level HLM models controlling for baseline scores, gender, age, and maternal education, children's classroom interaction in the Fall pre-school year were related to changes in social and self-regulatory skills during the pre-school year. Children's

competent task-oriented and conflictual interactions observed in the Fall were related to changes in both teacher-reported and directly assessed self-regulation during the pre-school year (Downer, Booren, Luckner, & Pianta, 2009).

Construct Validity

The inCLASS dimensions were derived from an extensive literature review of the social, emotional, and academic behaviors which develop during the pre-school years. It is therefore expected that there will be age differences in children's observed competencies and behaviors. The inCLASS observations were averaged across data collection time points to create a single score. In order to establish construct validity, a multivariate effect for age (calculated at the end of the year) was examined, $F(18, 304) = 3.71, p \leq 0.001$, followed by univariate ANOVAs for interpretation. Results indicate that inCLASS observations are somewhat sensitive to age differences, providing initial evidence of construct validity. As expected, 4-year-olds scored higher than 3-year-olds on peer sociability, peer assertiveness, peer communication, task engagement, and self-reliance, while 5-year-olds scored higher than 4-year-olds on peer sociability, assertiveness, and communication. One pattern stood out from the previous trends: 3-year-olds were rated higher than 4-year-olds for positive engagement with teacher, perhaps reflecting a tendency for younger children to more closely orient their classroom experiences around the teacher (Downer et al., 2009a; Booren, Downer, Luckner, Lima, & Pianta, 2008).

Because one goal of the inCLASS is to identify individual differences in school readiness, it was also important to assess its sensitivity to gender across the sampled age range (3-5 year olds). There were no differences in inCLASS scores across boys and girls (Downer, et al., 2009a; Downer, Booren, Luckner, & Pianta, 2009). Additionally, construct validity related to income-to-needs was investigated in the field study. Findings suggest that children from lower income families were observed to interact less competently with peers and in tasks, thus extending achievement/behavioral gap literature on teacher-related and directed assessed outcomes to observed classroom interactions (Downer, Booren, Luckner, & Pianta, 2009).

Finally, the three domains of the inCLASS were developed using a theoretical framework, which has been tested in an exploratory factor analysis with data from the pilot study. Findings support the three domains of children's positive classroom interactions (teacher, peer and task), plus an additional negatively-toned domain encompassing teacher and peer conflict interactions (Downer et al., 2009a). All factor loadings were moderate to high, and each factor had adequate internal consistency (Downer et al., 2009a).

Comments

The inCLASS was previously titled the CLASS-C.

The U.S. Departments of Health and Human Services and Education are funding the development and validation of the inCLASS as part of the Interagency Consortium

for School Readiness Outcome Measures (ICSROM). There are additional plans in place to develop the following:

- Training materials, and procedures for on-line data entry;
- Individual child reports for teachers and administrators; and
- Protocols for utilizing the CLASS and the inCLASS for intervention purposes.

References and Additional Resources

- Booren, L. M., Abry, T., Luckner, A. E., Yoder, B., Lima, O. K., Downer, J. T., & Pianta, R. C. (2008, May). *Examining a preschool observational assessment: Associations with teacher ratings and predictive validity of the CLASS-C**. Poster presentation for the annual meetings of the Association of Psychological Sciences, Chicago, IL.
- Booren, L. M., Downer, J. T., Luckner, A. E., Lima, O. K., Pianta, R. C. (2008, June). *Exploring the CLASS-C*: Associations among children's age, observed classroom behaviors, and teacher ratings*. Poster presentation symposium for the biennial Head Start's National Research Conference, Washington, D.C.
- Downer, J. T., Booren, L. M., Lima, O. A., Luckner, A. E., & Pianta, R. C. (2009a). The Individualized Classroom Assessment Scoring System (inCLASS): Preliminary reliability and validity of a system for observing preschoolers' competence in classroom interactions. *Early Childhood Research Quarterly*.
- Downer, J. T., Booren, L. M., Lima, O. A., Luckner, A. E., & Pianta, R. C. (2009b). *The Individualized Classroom Assessment Scoring System (inCLASS) technical manual*. Unpublished report, University of Virginia.
- Downer, J. T., Booren, L. M., Luckner, A. E., & Pianta, R. C. (2009, April). *Psychometric Results from a Field Test of the Individualized Classroom Assessment Scoring System (inCLASS)*. Poster presentation symposium for the biennial meeting of the Society for Research in Child Development, Denver, CO.
- Downer, J. T., Luckner, A. E., Booren, L. M., Lima, O. K., & Yoder, B. (2008, June). *Multi-level modeling of observational ratings using the Classroom Assessment Scoring System – Child Version (CLASS-C*)*. Poster presentation accepted for the biennial Head Start's National Research Conference, Washington, D.C.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System*. Baltimore, MD: Brookes Publishing.

Infant/Toddler Environment Rating Scale Revised Edition (ITERS-R)

I. Background Information

Author/Source

Source: Harms, T., Cryer, D., Clifford, R. M. (2003). *Infant/Toddler Environment Rating Scale - Revised Edition*. New York, NY: Teachers College Press.

Publisher: Teacher's College Press
1234 Amsterdam Avenue
New York, NY 10027
Phone: 1-800-575-6566
Website: www.tcpres.com

Purpose of Measure

As described by the authors:

The ITERS-R measures global quality in center-based programs serving children from birth to 30 months of age. "The ITERS-R contains items to assess provision in the environment for the protection of children's health and safety, appropriate stimulation through language and activities, and warm, supportive interaction" (Harms, Cryer & Clifford, 2003; p. 1).

The ITERS-R is a revision of the ITERS originally published in 1990. "The ITERS-R retains the original broad definition of environment including organization of space, interaction, activities, schedule, and provisions for parents and staff" (Harms, Cryer & Clifford, 2003; p. 1).

A Spanish Language version of the ITERS-R is available from Teachers College Press (www.teacherscollegepress.com/assessment_materials.html). In addition, translations of the scale into a variety of other languages are available. Contact the authors (www.fpg.unc.edu/~ecers/) for more information.

Population Measure Developed With

The original ITERS was developed with typical child care programs in North Carolina, but later work and revisions have been based on data from a wide range of program types including Early Head Start programs, typical child care centers, and programs for children with special needs. Special efforts were made to build on input from and experience with programs serving diverse populations including variations in race and ethnicity, type of special need and levels of income. Revisions were based on extensive use of the ITERS in various parts of the US and in other countries.

Age Range/Setting Intended For

The ITERS-R is used in center-based classrooms serving children from birth to 30 months of age

Ways in which Measure Addresses Diversity

- Indoor Space (item #1) assesses whether the space is accessible to children and adults with disabilities.
- Room Arrangement (item #4) assesses whether spaces for play are accessible to children with disabilities.
- Using Books (item #14) assesses whether there are a variety of books about people of different races, ages and abilities
- Active Physical Play (item #16) assesses whether some equipment can be used by children with disabilities.
- Music and Movement (item #18) assesses whether music from different cultures and in different languages is represented
- Dramatic Play (item # 20) assesses whether props such as dolls and equipment are provided to represent diversity of cultures and abilities
- Promoting Acceptance of Diversity (item #24): assesses whether the materials and activities represent and portray positively different races, cultures, ages, gender and abilities
- Provision for Children with Disabilities (item #32): 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 39 items categorized into seven subscales. Items are scored on a 7-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* (5 items)
 - Indoor space
 - Furniture for routine care and play
 - Provision for relaxation and comfort
 - Room arrangement
 - Display for children
- *Personal Care Routines* (6 items)
 - Greeting/departing
 - Meals/snacks
 - Nap

- Diapering/toileting
- Health practices
- Safety practices
- *Listening and Talking* (3 items)
 - Helping children understand language
 - Helping children use language
 - Using books
- *Activities* (10 items)
 - Fine Motor
 - Active physical play
 - Art
 - Music and movement
 - Blocks
 - Dramatic play
 - Sand and water play
 - Nature/science
 - Use of TV, video, and/or computers
 - Promoting acceptance of diversity
- *Interaction* (4 items)
 - Supervision of play and learning
 - Peer interaction
 - Staff-child interaction
 - Discipline
- *Program Structure* (4 items)
 - Schedule
 - Free play
 - Group play activities
 - Provisions for children with disabilities
- *Parents and Staff* (7 items)
 - Provision for parents
 - Provisions for personal needs of staff
 - Provisions for professional needs of staff
 - Staff interaction and cooperation
 - Staff continuity
 - Supervision and evaluation of staff
 - Opportunities for professional growth

Comments

The ITERS-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 ITERS-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/>.

The ITERS-R and other ERS instruments are also available in electronic form for use on Tablet PC machines through software package developed by the Branagh Information Group (<http://www.ersdata.com>) under license from Teachers College Press. This package is most appropriate for medium and large scale users.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The ITERS-R should be used by a trained observer. The authors recommend that at least 2.5 to 3 hours be spent observing in the classroom and note that spending more than 3 hours observing is preferable. An additional 20 – 30 minutes is needed to ask the teacher questions to help score indicators that were not observed. The ITERS-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 ITERS-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~~ participate in a training sequence led by an experienced ITERS-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 led by an experienced group leader, followed by an interrater agreement comparison. Additional field practice observations may be needed to reach the desired level of agreement or to develop reliability within a group” (Harms et al., 2003, 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 ITERS-R* book (Riley, Harms, & Cryer, 2004) that offers detailed information and photos that 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 ITERS-R* will assist groups of ITERS-R observers in developing reliability and being more consistent with the ITERS-R authors.

Setting

Observations are made in classrooms within center-based settings. Infants/toddlers are observed in one room or one group at a time.

Time Needed and Cost

Time: Outside observers will require at least three hours to use the measure.

Cost: Infant/Toddler Environment Rating Scale 2006 \$19.95

Spanish ITERS-R 2004 \$19.95

Video Observations for the ITERS-R (Training package): \$59.00

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Reliability was established in a field test with six observers who conducted 45 paired observations in programs in North Carolina. Inter-rater reliability was calculated at the indicator level and at the item level on the scale.

Indicators:

Average agreement on the 467 indicators on 39 items in the ITERS-R was 91.65%. The developers also calculated agreement on items 1-32 since many researchers omit the Parents and Staff subscale. Average agreement on the 378 indicators from items 1-32 was 90.27%

Items:

Agreement within 1 point (on a seven point scale) across the 39 items ranged from 64% on Item 4 (Room Arrangement) to 98% for Item 38 (Evaluation of Staff). Agreement on the 32 child-related items (omitting the Parents and Staff subscale) was 83%.

Cohen's Kappa was also computed. Across the 39 items, the weighted Kappa was .58. Across the 32 child-related items, the weighted Kappa was .55. The authors examined all items with a weighted Kappa below .50 and made minor changes to improve reliability.

Intraclass Correlations

The intraclass correlation was .92 for the 39 items and also for the 32 child-related items.

Intraclass correlations for the subscales were as follows:

- *Space and Furnishings* = 0.73
- *Personal Care Routines* = 0.67
- *Listening and Talking* = 0.77
- *Activities* = 0.91
- *Interaction* = 0.78
- *Program Structure* = 0.87
- *Parents and Staff* = 0.92

Internal Consistency

Cronbach's alpha was .93 for the 39 items and .92 for the 32 child-related items.

Alphas for the subscales were as follows:

- *Space and Furnishings* = 0.47
- *Personal Care Routines* = 0.56
- *Listening and Talking* = 0.79
- *Activities* = 0.79
- *Interaction* = 0.80

- *Program Structure* = 0.70
- *Parents and Staff* = 0.68

Because of alphas below .6 on the Space and Furnishings and Personal Care Routines subscales, the authors recommend that those subscales be used with caution. They also note that Item 32 (Provisions for Children with Disabilities) should be excluded from the Program Structure subscale unless most programs in the sample include children with disabilities.

Validity Information

Concurrent Validity

The authors cite studies showing that ECERS and ITERS scores are predicted by structural measures of quality such as child-staff ratios, group size, and staff education levels” (Cryer, Tietze, Burchinal, Leal, & Palacios, 1999; Phillipsen, Burchinal, Howes, & Cryer, 1998). The scores are also related to other characteristics normally expected to be related to quality such as teacher salaries and total program costs (Cryer et al., 1999; Marshall, Creps, Burstein, Glantz, Robeson, & Barnett, 2001; Phillipsen et al., 1998; Whitebook, Howes, & Phillips, 1989)” (Harms et al., 2003, p. 2).

Predictive Validity

The authors cite two studies showing that children’s development is predicted by scores on the rating scales (Burchinal, Roberts, Nabors, & Bryant, 1996; Peisner-Feinberg et al., 1999). The authors note that —the concurrent and predictive validity of the original ITERS is well established and the current revision maintains the basic properties of the original instrument” (Harms et al., 2003, p. 2).

References and Additional Resources

- Burchinal, M. R., Roberts, J. E., Nabors, L. A., & Bryant, D. M. (1996). Quality of center child care and infant cognitive and language development. *Child Development, 67*, 606-620.
- Cryer, D., Tietze, W., Burchinal, M., Leal, T., & Palacios, J. (1999). Predicting process quality from structural quality in preschool programs: A cross-country comparison. *Early Childhood Research Quarterly, 14*, 339-361.
- Harms, T., Cryer, D., Clifford, R. M. (2003). *Infant/Toddler Environment Rating Scale - Revised Edition*. New York, NY: Teachers College Press.
- Marshall, N. L., Creps, C. L., Burstein, N. R., Glantz, F. B., Robeson, W. W., & Barnett, S. (2001). *The Cost and Quality of Full Day, Year-round Early Care and Education in Massachusetts: Preschool Classrooms*. Wellesley, MA: Wellesley Centers for Women and Abt Associates, Inc.
- Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., Yazejian, N., Byler, P., Rustici, J., & Zelazo, J. (1999). *The children*

- of the cost, quality, and outcomes study go to school: Technical report.* Chapel Hill: University of North Carolina at Chapel Hill, Frank Porter Graham Child Development Center.
- Phillipsen, L., Burchinal, M., Howes, C., & Cryer, D. (1998). The prediction of process quality from structural features of child care. *Early Childhood Research Quarterly, 12*, 281-303.
- Riley, C., Harms, T. & Cryer, D. (2004). *All about the ITERS-R. A detailed guide in words and pictures to be used with the ITERS-R.* PACT House Publishing.
- Whitebook, M., Howes, C., & Phillips, D. (1989). *Who cares? Child care teachers and the quality of care in America. National child care staffing study.* Oakland, CA: Child Care Employee Project.

Language Interaction Snapshot (LISn)

I. Background Information

Author/Source

Source: Atkins-Burnett, S., Sprachman, S., & Caspe, M. (2010). *Language Interaction Snapshot (LISn)*. Princeton, NJ: Mathematica Policy Research.

Sprachman, S., Caspe, M. & Atkins-Burnett, S. (2009). *Language Interaction Snapshot (LISn) Field Procedures and Coding Guide*. Princeton, NJ: Mathematica Policy Research.

Publisher: Please contact Susan Sprachman at:
SSprachman@mathematica-mpr.com

Purpose of Measure

As described by the authors:

The LISn is designed to examine how the language environment differs for children, particularly in classrooms that include dual language learners. It focuses on individual children and the language provided to each child by the lead teacher and other adults in the classroom as well as the language used by each of the selected children. Since most early childhood classrooms spend a limited amount of time in large group, the majority of the interactions are not shared by all the children in the class. The LISn allows examination of the interactions experienced by individual children which can then be aggregated to the group or classroom level. In addition, end of visit ratings provide information about the social climate and the instructional supports for language.

Population Measure Developed With

Initial pre-testing of the LISn took place in one bilingual early childhood classroom in New Jersey. Two hours of morning activities were videotaped and later coded by trained LISn observers. A 2007 pilot study of the LISn measure consisted of observations in 18 classrooms that were part of the Los Angeles Universal Preschool program and 26 classrooms that are part of San Francisco's Preschool for All program, both of which have large populations of English language learners. Three children in each classroom were selected as focal children for observations using stratified random sampling. Each child was observed 3 times. In fall 2009 a pilot study with 18 classrooms and 4 family child care providers was conducted collecting 6 observations per child during half-day session.

Age Range/Setting Intended For

The LISn can be used to assess early childhood center-based settings as well as family child care settings. The LISn is intended for use in classrooms with 3- to 4-year-old children.

Ways in which Measure Addresses Diversity

A main focus of the LISn is to record child and teacher verbal communication by examining the experience of one focus child throughout the observation. The language used by both the child and teacher is documented throughout the observation. The measure is intended for use in classrooms where dual language learners are present. Observers code whether English or a language other than English is being spoken.

Key Constructs & Scoring of Measure

The LISn documents early childhood classroom language environments during each 5-minute coding period for a focal child (Atkins-Burnett et al., 2009a; 2009b) including the following areas:

- *Language Spoken*: This code is used when a focus child or teacher in the classroom speaks (English, another language, or in mixed utterances).
- *Focus Child Verbal Communication*: This code identifies the focus child's conversational partner
- *Type of Teacher Verbal Communication*: This code is used to identify the purpose of the teacher's verbal communication with a focus child (e.g., repeats or confirms, requests language, gives directions, etc.).
- *Global Classroom Setting*: This captures the content (e.g., math, singing, science) and structure (e.g., small group, whole group, free choice/centers) during that 5-minute observation period referred to as a snapshot. It also indicates whether sustained conversations occurred with different conversational partners.

Each category of language is coded as present or absent for each of ten 30-second cycles within each five minute snapshot. The number of times that each category is observed is summed across the snapshot and then a sum or mean score is created for each category (by language).

Preliminary analysis indicates that several factors can be obtained from the LISn snapshots at the child level. With six snapshots per child, estimates of different types of adult language can be constructed:

- Child to child talk
- Child to teacher talk
- Responsive teacher language: repeating or confirming child language, elaborating on child language
- Teacher instruction using contextualized language
- Teacher instruction that includes responsive language and decontextualized language
- Sustained conversations with other children

- Sustained conversations with teachers

The newly developed End of Visit Ratings for the LISn are hypothesized to provide estimates of different dimensions of the instructional supports offered for language development and the social climate (Atkins-Burnett et al, 2010):

- Intentional instruction of vocabulary and language
- Sheltered English approaches to supporting understanding (e.g., use of visual supports, emphasis of key words, repeating)
- Behavior management
- Active child engagement
- Positive social relationships
- Productivity

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Trained field observers administer the measure.

Training Required: Observers are trained over a two-day period with 4 hours of self study prior to training which includes reading the coding manual and completing practice coding for three scenarios. The first day of training consists of presentations, discussion, and practice coding from videotapes. The second day consists of a series of field practices and discussions and the establishing of reliability.

Setting

Observations are made in family and center-based child care setting serving 3- and 4-year-olds.

Time Needed and Cost

Time: A time sampling method is used for the LISn. During 5-minute intervals, observers observe in 30 second segments, totaling 10 segments per 5-minute period. In the pilot study, each child was observed for three 5-minute periods.

Cost: There is no cost for the use of the measure however at least one trainer from each user needs to attend a training conducted by Mathematica. The cost of the training varies depending upon how many people are trained at the time.

III. Functioning of Measure

Reliability Information

In pilot testing the LISn measure, observers established inter-rater reliability with a lead trainer. The developers calculated video inter-rater reliability and field inter-rater reliability on the overall codes and the teacher and child components individually. The inter-rater agreement on the video reliability for overall coding was 96%. Across studies, codes and language, inter-rater agreement ranged from 72 to 99% on individual codes, with overall agreement of 89%.

Validity Information

Criterion Validity

In the 2007 pilot study in Los Angeles, there was not enough talk observed in "Spanish or Other language" during this end of the year observation. Only the English variables had enough variance to construct reliable scales. The "Lead Teacher Talk in English" ($\alpha = 0.77$) and "Other Adult Talk in English" ($\alpha = 0.72$) included all the categories of talk except singing and elaboration. Singing was not related to the other categories of talk, and elaboration of child language was noted only four times (once per child). After the first pilot, the observer training about when to code "elaboration" was strengthened to assure that the low incidence is real and not a function of poor observer training. Psychometric information from the second pilot is not yet available.

Concurrent Validity

Initial evidence of validity was obtained by examining relationships with the observations of the same classroom using the CLASS (Pianta, LaParo, and Hamre 2008). In the Los Angeles sample of 14 classrooms with complete data, moderate correlations were found between CLASS Instructional Support and Total Talk in English ($r = .55$) and Gives Information in Context ($r = .63$). The dimensions of the CLASS Instructional Support scale were not all related to the LISn variables. The strongest relationships were found between the LISn variables about requests for language, giving both contextualized and decontextualized information, repeating or confirming child language, and the total talk with the CLASS Quality of Feedback ($r = .64$ to $r = .72$). LISn variables for elaborating, requesting language from children, and giving information showed moderate relationships with CLASS Language Modeling ($r = .52$ to $.69$). The LISn variables showed no correlation with the CLASS dimension on Concept Development.

In the pilot sample in San Francisco that used the LISn and the CLASS, these relationships were not detected. The sample of San Francisco classrooms seldom used large or small group instruction. It is likely that in classrooms where instruction is more individualized, the time sampling of interaction with children does not represent well what the teacher does, though it may represent the experience of individual children. In those classroom, more time samples may be needed per child in order to obtain a valid picture of the interactions. The relationships were based on only 3 five-minute observations per child for 3 children.

Predictive Validity

There have been no tests for predictive validity at this time.

Construct Validity

The LISn was developed based on a review of available measures and literature on language and literacy development among dual language learners. The developers sought input from a group of experts on the items included and reviewed items in the Child Classroom Observation System (C-COS; Boller, & Sprachman, 1998) and the Child Care Assessment Tool for Relatives (CCAT-R; Porter, Rice, & Rivera, 2006).

References and Additional Resources

- Atkins-Burnett, S., Sprachman, S., & Caspe, M. (2009a). *Capturing quality in dual language learner classrooms*. Presentation to the biannual meeting of the Society for Research in Child Development, Denver, CO.
- Atkins-Burnett, S., Sprachman, S., & Caspe, M. (2009b). Descriptive Results From the Pilot Study of the Language Interaction Snapshot (LISn), Mathematica Policy Research, Princeton, NJ.
- Atkins-Burnett, S., Sprachman, S., & Caspe, M. (2010). *The Language Interaction Snapshot (LISn)*. Mathematica Policy Research, Princeton, NJ.
- Boller, K., & Sprachman, S. and the Early Head Start Research Consortium (1998). *The Child-Caregiver Observation System Instructor's Manual*. Mathematica Policy Research, Princeton, NJ.
- Love, J. M. Atkins-Burnett, S., Vogel, C., Aikens, N., Xue, Y., Mabutas, M., Carlson, B. L., Sama Martin, E., Paxton, N., Caspe, M., Sprachman, S., & Sonnenfeld, K. (2009). Los Angeles Universal Preschool Programs, Children Served, and Children's Progress in the Preschool Year: Final Report of the First 5 LA Universal Preschool Child Outcomes Study, Final Report Mathematica Policy Research, Princeton, NJ. (<http://www.first5la.org/files/UPCOS-Final-Report.pdf>)
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System (CLASS) Manual*. Charlottesville, VA: Center for Advanced Study of Teaching and Learning.
- Porter, T., Rice, R., & Rivera, E. (2006). Assessing quality in family, friend and neighbor care: The child care assessment tool for relatives. New York, NY: Institute for a Child Care Continuum.
- Sprachman, S., Caspe, M., Atkins-Burnett, S., López, M., Parrish, D., Shkolnik, J., et al., (2009). *The Language Interaction Snapshot (LISn): Examining Language Interactions in Linguistically Diverse Classrooms*. Presentation to the biannual meeting of the Society for Research in Child Development, Denver, CO.

Observation Measures of Language and Literacy (OMLIT)

I. Background Information

Author/Source

Source: Goodson, B. D., Layzer, C. J., Smith, W. C., & Rimdzius, T. (2006). *Observation Measures of Language and Literacy Instruction in Early Childhood* (OMLIT). Cambridge, MA: Abt Associates, Inc.

Publisher: Abt Associates, Inc.
55 Wheeler Street
Cambridge, MA
<http://www.abtassociates.com/>

Purpose of Measure

As described by the authors:

The Observation Measures of Language and Literacy Instruction in Early Childhood Education Classrooms (OMLIT) was developed as a battery of measures “to address the need for research-based, reliable and valid measures of the instructional practices and environmental supports for language and literacy in early childhood classrooms” (Abt Associates, undated, p. 1). The OMLIT includes six instruments (Classroom Description, Snapshot of Classroom Activities (SNAPSHOT), Read-Aloud Profile (RAP), Classroom Literacy Instruction Profile (CLIP), Quality Rating of Language and Literacy Instruction (QUILL), and the Classroom Literacy Opportunities Checklist (CLOC). The Arnett Caregiver Rating Scale (Arnett, 1989) is typically done along with the OMLIT measures. Although individual OMLIT measures may be used alone, together the measures provide an in-depth assessment of the quality (and in some cases quantity) of the language and literacy activities in the classroom.

The OMLIT was designed as a research tool. The first version was developed based on findings from a national conference on instructional practices related to early literacy (Abt Associates, 2003). The OMLIT was then refined and adapted for the Even Start Classroom Literacy Intervention and Outcomes Study (CLIO) under contract ED-01-CO-0120, as administered by the Institute of Education Sciences. It has since been used in several other experimental studies that are evaluating language/literacy curricula. In addition, the OMLIT-Snapshot (adapted) and the OMLIT-RAP are being used in a randomized experiment conducted in family child care homes in Massachusetts.

Population Measure Developed With

- Early versions of the OMLIT were piloted in Fall 2003 in three child care centers in the Boston area by six observers. These centers were all day programs serving primarily subsidized children.

- A total of 16 observers were trained on a revised OMLIT in spring 2004. The OMLIT was used in 2004 in the national sample of Even Start programs in the CLIO study and a large sample of child care centers serving subsidized children in Miami-Dade county (Project UPGRADE).
- The OMLIT was further revised after 2004. The revised version was used in the national sample of Even Start programs in the CLIO study, in the same sample of child care programs in Miami-Dade county, and in a sample of public school pre-kindergarten programs run by the Chicago School Department.
- Final revisions involved primarily formatting and additional definitions of codes. The most current version reviewed is dated February 2006.

Age Range/Setting Intended For

The OMLIT was developed for observing early childhood classrooms. In addition, the Read Aloud Profile has been used in family child care homes.

Ways in which Measure Addresses Diversity

Research on the acquisition of English in English language learners informed the development of the OMLIT. One of the OMLIT measures, the Classroom Literacy Opportunities Checklist (OMLIT-CLOC), includes a question on whether there is cultural diversity in literacy materials. The OMLIT-Snapshot (a description of classroom activities and groupings) includes a question on whether adults and children are speaking in English or another language. Finally, the OMLIT-QUILL is used to assess the overall quality of instructional practices in language and literacy with English language learners.

Key Constructs & Scoring of Measure

The OMLIT is made up of six separate measures and the Arnett Caregiver Rating Scale (Arnett, 1989).

- The *Classroom Description* (OMLIT-Description) has six sections. Four are completed at the beginning of the observation and require the observer to ask a few questions of the teacher:
 - *Setting Profile*. Includes address of setting, name of setting, date of observation, start and end times for observation, and observer name.
 - *Staff*. Includes listing of all classroom staff present and assigns unique IDs to each staff member.
 - *Child Population*. Includes number of children by age and by home language, presence of any children with diagnosed special needs.
 - *Classroom Theme*. Includes any current classroom theme(s) identified by the teacher.

Two other sections of the *Classroom Description* are completed at the end of the observation.

- *Language(s) of Instruction*. For each member of the classroom staff, the observer indicates the proportion of time English, Spanish, or another language was used during instruction with the children. The observer also

indicates whether there was at least one adult in the classroom who spoke the language of every child.

- *Atypical Observation*. The observer indicates whether there was something about the observation that made it atypical for that classroom, such as an emergency in the classroom, an extended fire drill, etc.
- The *Snapshot of Classroom Activities* (OMLIT-Snapshot) has two sections:
 - *Environment*. Includes codes for total number of children present, total number of adults present, and type of adults present (teachers, aides, other adults). This section can be used to compute staff/child ratio. It also includes a check box to indicate that during this Snapshot, all of the children are doing the same activity.
 - *Activities*. Includes codes for type of activity; number of children, teachers, aides, and other adults in the activity, number of other adults in activity; integration of print in activity; and language(s) spoken by children or adults, if any. These individual activity codes can be combined to form activity constructs, such as early literacy activities or developmental activities; child grouping constructs, such as individual child activities, pairs, small groups, medium groups, and large groups of children; and teacher engagement constructs.
- The *Read Aloud Profile* (OMLIT-RAP) has seven sections.
 - *Pre-reading* (11 items)
 - *Reading* (14 items)
 - *Post-reading* (11 items)
 - *Adult reading book* (teacher, assistant, other adult)
 - *Adult language with children* (English, Spanish, other)
 - *Number of children reading*
 - *Book characteristics* (6 items)

These RAP codes have been used to form constructs for support of comprehension (providing information about text, introducing new vocabulary, asking questions, reviewing text and providing extension activities), support of print motivation, and support of phonological awareness/print knowledge.

In addition, the OMLIT-RAP includes coding of three features of the read aloud on a 1 (minimal) to 5 (high) scale:

- *Story-related vocabulary*
- *Adult use of open-ended questions*
- *Depth of post-reading*
- The *Classroom Literacy Opportunities Checklist* (OMLIT-CLOC) is an inventory of classroom literacy resources. It identifies 11 aspects of the literacy environment, each of which is rated on a 1 (minimal) to 3 (high) scale:

- *Physical layout of the classroom* (5 items)
 - *Text or print environment* (8 items)
 - *Literacy-related materials and toys* (2 items)
 - *Books and reading area* (12 items)
 - *Listening area* (3 items)
 - *Writing supports* (6 items)
 - *Literacy Materials Outside of the Reading and Writing Areas* (3 items)
 - *Diversity in literacy materials* (3 items)
 - *Instructional technology* (2 items)
 - *Richness of curriculum theme and integration of theme in classroom activities, materials, displays* (7 items)
 - *Literacy resources outside of the classroom* (4 items)
- The *Classroom Literacy Instruction Profile* (OMLIT-CLIP) involves a two-stage coding protocol. First, the observer determines if any classroom staff member is involved in a literacy activity. If so, the observer codes seven characteristics of the literacy activity:
- *Type of activity*
 - *Literacy knowledge being afforded to the children*
 - *Teacher's instructional style*
 - *Text support/Context for literacy instruction*
 - *Number of children involved in activity with teacher*
 - *Languages spoken by staff and children, and focus of the language* (i.e., talk with peers, talk with group, talk with individual children, etc.)
- If the literacy activity involves adult-child discussion, the quality of the discussion is evaluated on a 1 (minimal) to 5 (extensive) scale for two characteristics:
- *Cognitive challenge in the discussion* (3 items)
 - *Depth of the discussion* (2 items)
- The *Quality Rating of Language and Literacy Instruction* (OMLIT-QUILL) is an overall evaluation of the quality and quantity of the instructional practices around literacy. Ten items are coded for frequency (no opportunities, one, a few, or many) and also are rated on a 1 (minimal) to 5 (high) scale for overall quality (with examples offered for anchors at the 1, 3, and 5 rating levels). The ten items address:
- *Opportunities to engage in writing*
 - *Attention to/promotion of letter/word knowledge*
 - *Opportunities/encouragement of oral language to communicate ideas and thoughts*
 - *Attention to the functions and features of print*
 - *Attention to sounds in words throughout the day*
 - *Attention to/promotion of print motivation*
 - *English Language Learner (ELL) children intentionally included in activities, conversations*

- *Development of both home language(s) and English supported for ELL children*
- *Home language(s) of ELL children integrated into language and literacy activities*
- *Language and literacy materials/methods appropriate for ELL children*
- *Opportunities for dramatic play and play planning*
- *Integration of special needs children in classroom*

In addition, a total rating of support for language and literacy is computed as the average of the first five items.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Trained observers may use the battery of measures.

To date, the OMLIT has been administered in the following way. The Classroom Description is completed at the very beginning of the observation. Subsequently, a Snapshot is completed every ten minutes as an instantaneous picture of the activities and groupings in the class at the ten-minute mark. The CLIP is completed every ten minutes as well, but starting on the five-minute mark within the Snapshot. The CLIP involves watching the teacher and aide for five minutes, and then coding any literacy activity in which they are involved over the next five minute period.

The RAP is completed at any time during the observation that a target adult in the classroom begins to read aloud to a designated number of children (the number could vary by study, e.g., in the CLIO study, the RAP was only recorded if the teacher or aide read to at least two children). While it is being coded, the observer does not complete a Snapshot or CLIP, although when the RAP is completed, the observer goes back and indicates that reading aloud was occurring in any coding interval covered by the RAP.

The QUILL is completed at the end of the observation, and is based on evidence from all of the other measures, as well as notes on events that occurred outside of the coding windows. The CLOC is also completed at the end of the observation.

Training Required: —The amount of training required depends on how many and which of the separate OMLIT measures are being used. The measures require two types of training: (a) classroom training, culminating in paper-and-pencil reliability tests, and (b) practice observation in a pre-school classroom. Ideally, inter-rater reliability is also assessed for each trained observer through dual (simultaneous) coding in a pre-school classroom by the observer and an OMLIT trainer. The four central measures (Snapshot, RAP, CLIP, QUILL) require 8 hours each of classroom training. The OMLIT-CLOC and the Classroom Description require less than one-half day of classroom training (Layzer, 2006, personal communication).

"The cost of training depends on the number of measures to be trained. Minimally, each of the central measures requires one day of training by one trainer for up to 10 trainees at \$1,000 a day plus expenses. Two trainers are required for more than 10 trainees" (Layzer, 2006, personal communication).

Setting

The OMLIT was developed for observations in early childhood education classrooms, although the Snapshot and RAP are currently being used in family child care homes as well.

Time Needed and Cost

Time: The authors recommend observing for a minimum of 3 hours (approximately one half-day) in the classroom.

Cost: A PDF copy of each OMLIT measure may be obtained from Abt Associates, Inc. A training manual is also available electronically.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Paired observers coded classroom activities as part of the training process (14 paired observations) as well as during actual data collection for the Even Start Classroom Literacy Interventions and Outcomes Study (17 paired observations). Inter-rater reliability calculations are based on data from both of these sources (Goodson & Layzer, 2005).

- OMLIT-Snapshot: "High inter-rater agreement was not expected for many of the Snapshot codes, since the allocation of children to activities could vary depending on the direction of rotation of the observer's scan of the classroom. For this reason, while we expected that observers might agree on the activities taking place in the classroom, they were much more likely to differ on the number of children they assigned to each activity. This also leads us to believe that the inter-rater reliability estimates for the Snapshot present an underestimate of the true level of agreement across trained observers in how they would code an idealized 'stationary' classroom.

The Environment section of the Snapshot includes a count of the numbers of children and adults present in the classroom. There was a high level of agreement – above 80% - on all codes on the Environment section. On the Activities section of the Snapshot, children and adults are allocated into activities. This is the part of the Snapshot where small differences in timing between observers could adversely affect their agreement. As predicted, the inter-rater agreement was lowest for the categories involving numbers of children in an activity (57%). The level of agreement on the numbers of adults in each activity also was low. On the other hand, the types of activities

that each observer coded had higher inter-rater agreement (82%), as did the integration of literacy in activities (88%). Although the level of agreement at the activity level on whether or not children or adults were talking was only 71%, agreement was very high – 100% - on whether or not there were any adults or children talking in *any* of the activities coded on the Snapshot" (Abt. Associates, Attachment B, 2007, p. B7-B8).

- OMLIT-RAP: Inter-rater agreement on strategies used before, during, and after a read aloud ranged from 85% to 97%, with an overall average of 90%. The inter-rater agreement on individual instructional codes during reading ranged from 53% to 93%. The average agreement on the Quality Indicators was high if agreement was defined as within one point (83% for story-related vocabulary, 83% for adult use of open-ended questions, 85% for depth of post-reading activity, and 84% across all quality indicators).

However, if agreement was defined as exactly the same quality rating across observers, the percent agreement dropped substantially (76% for story-related vocabulary, 64% for adult use of open-ended questions, and 76% for depth of post-reading activity).

- OMLIT-CLOC: It is reported that nine of the ten sections of the CLOC had reliabilities above 70% (three sections had agreement above 80%: writing resources 81%, literacy toys and materials 82%, and physical layout of classrooms 91%).

Researchers at Abt Associates indicated that they would ~~–~~strive to increase the reliability of this section through (a) improving the definition of the item to help observers understand what they are looking for, and (b) focusing training on these items to heighten observer awareness of isolated materials in different areas of the classroom" (Abt Associates, 2007, B-5).

- OMLIT-CLIP: Inter-rater agreement is based only on the 17 Even Start classrooms. The CLIP involves a two-stage coding process. Observers first determine whether any classroom staff are involved in a literacy activity. If so, then that activity is coded for additional information about the characteristics of the literacy activity.

On average, the inter-rater agreement on whether a literacy activity occurred was 85% (range of agreement across pairs: 50% to 100%). When both observers identified a literacy activity, there was high agreement on the characteristics of that activity (agreement ranged from 96% to 98% across the 7 characteristics). In addition, inter-rater agreement on the quality ratings in the CLIP were 92% for *cognitive challenge in the discussion* and 93% for *depth of the discussion*.

- OMLIT-QUILL: Inter-rater agreement is reported on six of the OMLIT-QUILL literacy activities, but excludes the four items concerning activities for ELL children.

Inter-rater agreement on the frequency of the six literacy activities ranged from 67% to 83%, with average overall agreement for frequency being 76%. Inter-rater agreement on the quality of literacy activities (within one point) ranged from 68% to 94% across the six literacy activities.

Validity Information

Content Validity

The measures were derived from information discussed at a research conference of experts in the field. In fall 2003, Abt Associates convened a conference on measuring the quality of language and literacy instruction in early childhood programs. The conference focused on research evidence of instructional practices linked to short or long-term outcomes for children. The OMLIT was developed around this research on instructional practices.

Comments

The reliability reported in Abt Associates Attachment B (2007) appears to be on the 2004 version of the OMLIT, and consequently does not reflect the current (February 2006) version of the OMLIT-CLOC.

The Abt Associates (undated) report that "the QUILL ratings and CLOC constructs have undergone IRT scaling by Futoshi Yamoto, a psychometrician at Abt, which shows these constructs to have very high reliability. A separate technical report has been prepared on the IRT scaling, and this will be available soon" (Abt Associates, 2007, p. B-9).

References and Additional Resources

Abt Associates, Inc. (2003). *Assessing instructional practices in early literacy and numeracy*. A conference held at Abt Associates, Inc, Cambridge, MA, 2002.

Abt Associates, Inc. (2007) *Attachment B in Evaluation of Child care Subsidy Strategies Findings from Project Upgrade in Miami-Dade County*. Abt Associates, Inc., Cambridge, MA.

Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10, 541-522.

Goodson, B. D. and Layzer, C. (2005). *Assessing support for language and literacy in early childhood classrooms: The Observation Measures for Language and Literacy (OMLIT: Goodson, Layzer, Smith, and Rimdzius, 2004)*. Paper presented at the annual conference of the American Educational Research Association, Montreal, Canada.

Goodson, B. D., Layzer, C. J., Smith, W. C., & Rimdzius, T. (2006). *Observation Measures of Language and Literacy Instruction in Early Childhood* (OMLIT). Cambridge, MA: Abt Associates, Inc.

Observational Record of the Caregiving Environment (ORCE)

I. Background Information

Author/Source

Source: NICHD Study of Early Child Care and Youth Development Phase I Instrument Document (<http://secc.rti.org/instdoc.doc>).
Observational Record of the Caregiving Environment (ORCE):
Behavior Scales, Qualitative Scales, and Observed Structural
Variables for 6, 15, 24, & 36 months.

NICHD Study of Early Child Care and Youth Development Phase II Instrument Document
(<http://secc.rti.org/Phase2InstrumentDoc.pdf>).
Observational Record of the Caregiving Environment (ORCE):
Behavior Scales, Qualitative Scales, and Observed Structural
Variables for 54 months.

NICHD Early Child Care Research Network (1996).
Characteristics of infant child care: Factors contributing to positive
caregiving. *Early Childhood Research Quarterly*, 11, 269-306.

NICHD Early Child Care Research Network. (2001). Nonmaternal
care and family factors in early development: An overview of the
NICHD Study of Early Child Care. *Journal of Applied
Developmental Psychology*, 22, 457-492.

Publisher: Ablex Publishing

Purpose of Measure

As described by the authors:

The Observational Record of the Caregiving Environment (ORCE) was created for the NICHD Study of Early Child Care (now known as the NICHD Study of Early Child Care and Youth Development) because no other observational rating scale had been developed which could address children's behavior over the entire age span of the study (6 months to 54 months) and across different non-maternal child care settings.

Although several other measures were sources of "inspiration" for the ORCE (i.e., CIS, Arnett, 1989; Assessment Profile for Early Childhood Programs, Abbott-Shim & Sibley, 1987; FDCRS and ECERS, Harms & Clifford, 1989, 1980), "the results of extensive piloting and much input from the Steering Committee as well as members of the child care subcommittee have made this an original and unique assessment instrument specifically designed for our purposes" (NICHD Study of Early Child

Care Phase I Instrument Document, 2004, p. 127). The ORCE was created "(a) to assess minute-to-minute evidence of caregiving and quality in a relatively objective, quantitative and qualitative way and (b) to accommodate to the demands of the enterprise and the limitations of the human observers (i.e., we tried to get as much detail and 'richness' as our coders could record reliably)" (NICHD Study of Early Child Care Phase I Instrument Document, 2004, p. 128). In contrast to other instruments which focus on aspects of quality in the classroom at large (e.g., ITERS, ECERS), the ORCE focuses on the proximal experiences of the child while in non-maternal care and provides information about (1) the behaviors of the caregivers toward the target child and (2) the behavior of the target child. The ORCE also provides observed data on "structural measures" such as the number of children in the group and the child to adult ratio.

Population Measure Developed With

The ORCE was developed based on 1,364 families across 10 sites nationally (NICHD Study of Early Child Care and Youth Development). The children were born in 24 hospitals and were followed up wherever they lived subsequently.

The ten research sites were in geographical proximity to the following universities or research organizations: University of Arkansas, Little Rock, AK; University of California at Irvine, Orange County, CA; University of Kansas, Lawrence, KS; Wellesley College, MA; Temple University, Philadelphia, PA; University of Pittsburgh, PA; University of Virginia, Charlottesville, VA; Western Carolina Center, Morganton and Hickory, NC; University of Washington, Seattle, WA; and University of Wisconsin, Madison, WI.

Age Range/Setting Intended For

The ORCE may be used for observational assessments of the non-maternal child care setting when the child is 6, 15, 24, 36 and 54 months old.

Ways in which Measure Addresses Diversity

Information regarding diversity with the ORCE was not available in materials reviewed.

Key Constructs & Scoring of Measure

The instrument has four versions, one for each time point when data were collected (the 24 & 36 month versions are the same). Each version contains three parts: (1) ORCE Behavior Scales, (2) ORCE Qualitative Ratings, and (3) ORCE Observed Structural Variables. The 6 month ORCE also contains Global ratings. Additionally, several composites were created for each version of the ORCE. For more information about composites, please refer to the Instrument Documents and Manuals (<http://secc.rti.org/>).

Behavior Scales

The Behavior Scales provide an account of the occurrence of specific behaviors directed by caregivers toward the target child. A behavior is either marked as having

occurred within a 30-second observation interval or left blank. Behaviors included in the scales were derived from the research on parental and caregiver behaviors that have been found to be associated with positive child development. At 24, 36, and 54 months, specific child behaviors are also recorded (NICHD Study of Early Child Care Phase I Instrument Document, 2004). A more detailed description of each behavior code can be found in the corresponding Instrument Documents and Manuals (<http://secc.rti.org/>).

6 Month Behavior Scales

- *Positive and Negative Affect*
 - Responds to negative affect
 - Shared positive affect
 - Positive physical contact
- *Language Focused Interaction*
 - Responds to child's vocalizations
 - Reads aloud to child
 - Other talk to child
- *Stimulation*
 - Stimulates cognitive development
 - Stimulates social development
- *Behavior Management*
 - Facilitates child's behavior
 - Restricts child's activities
 - Restricts in a physical container
 - Speaks negatively to child
 - Uses negative physical actions
- *Child's Activity*
 - Physical care
 - Other activity with adult
 - Activity with child(ren) only
 - Solitary activity
 - Watching/unoccupied/transition
 - Watching TV
- *Child's Interaction with Other Children*
 - Positive/neutral interaction
 - Negative interaction

At the 6 month observation, the NICHD Early Child Care Research Network (1996, p. 278) created a composite of the Behavioral Scales:

- *Positive caregiving frequencies.* Sum of Positive Behavior (shared positive affect + positive physical contact), Responsive Behavior (responds to vocalization + facilitates infant behavior), and Stimulating Behavior (stimulates cognitive development + stimulates social development + asks question + other talk + reads).

15 Month Behavior Scales

The items on the 15 month scale are the same as those included at 6 months with the following exceptions:

- *Adult Language* (replaced Language Focused Interaction and includes additional items)
 - Speaks positively to child/ren
 - Speaks negatively to child/ren
 - Asks question to child/ren
 - Gives direction to child/ren
 - Other talk to group
- *Activity Setting* (category added)
 - With adult (and child/ren)
 - With child/ren only
 - Alone
- *Child's Self Assertion* (category added)
 - Says "no"/refuses
 - Acts defiant

The following composites were created with the 15-month behavioral variables:

- *Total Stimulation*: Stimulates cognitive development + Stimulates social development + Reads aloud to children + Asks question to child + Other talk to child
- *No Stimulation*: Solitary activity + Unoccupied + Restricted in a physical container + Watching TV + Other activity with adult (reflected) + Activity with children only (reflected)
- *Response to Distress*: Proportion of time adult responds to negative affect, Out of Child's total exhibited negative affect
- *Responsiveness*: Responds to child's vocalizations + Facilitates child's behavior
- *Negative Contact*: Restricts child's activities + Negative talk to child + Negative physical contact with child
- *Positive Contact*: Shared positive affect + Positive physical contact + Positive talk
- *Rate of Physical Care*: Proportion of physical care out of total time spent with adult
- *Child's Contact with Peers*: Activity with children only + Negative interactions with children + Positive/Neutral interactions with children
- *Total Adult Attention*: Sum of adult attention paid to child during all segments
- *Total Adult Talk*: Gives direction + Positive talk + Negative talk + Reads aloud + Asks questions + Other talk to child + Other talk to group
- *Group Interactions*: Rate of other talk to group as proportion of total other talk
- *15-month Behavioral Composite, Standardized* (M=0, sd =3): Positive affect + Positive talk + Positive physical contact + Responds to child's vocalizations +

Asks questions + Other talk to child + Reads aloud + Stimulates child's cognitive development + Stimulates child's social development + Facilitates child's behavior + Restricts child's activities (reflected) + Negative talk to child (reflected) + Negative physical contact with child (reflected) + Child's time spent unoccupied (reflected)

24 and 36 Month Behavior Scales

The items on the 24 & 36 month scales are the same as those included at 15 months with the following exceptions:

- *Stimulation* (category added)
 - Teaches academic skill
 - Teaches social rule
 - Positive physical contact
 - Mutual exchange
- *Physical Control* (category added)
 - Negative/restricting actions
- *Child's Interaction with Other Children* (same category, item added)
 - Mutual pretend play
- *Child's Behavior* (category added)
 - Prosocial act
 - Negative act (nonaggressive)
 - Verbal aggression
 - Physical aggression
 - Complies with adult
 - Says —no/refuses to adult
 - Acts defiant to adult

54 Month Behavior Scales

The items on the 54 month scale are the same as those included at 24 and 36 months with the following exceptions:

- *Adult Language/Management* (this category includes all of the items in the Language scale with items added)
 - Answer's child's question
 - Offers choice
 - Negative management
- *Stimulation by Adult* (same category, item added)
 - Facilitates learning
- *Positive/Neutral Peer Activities* (new category)
 - Cooperative play
 - Boisterous play
 - Other positive/neutral interaction
 - Parallel play
- *Negative Peer Activities* (new category, but includes some of the same items in the Child's Behavior category from 24/36 mo. measure)

- Peer negative behavior
- Child physical aggression
- Child verbal aggression
- Child negative act (nonaggressive)
- *Child Alone* (new category, but includes some of same items in the Child's Activity category from 24/36 mo. measure)
 - Solitary activity
 - Watching/unoccupied/transition
 - Watching TV

According to the NICHD Early Child Care Study Phase II Instrument Document, at the 54 month observation, the Behavioral Scales generated the following composites (items within composites were first standardized with mean of 0 and sd of 1):

- *Positive caregiving*. Sum of Encourages or praises, Offers choice, Asks question, Gives direction, Adult other talk, Teachers academic skill, Facilitates learning and Playful exchange.
- *Peer agonism*. Sum of Peer negative behavior, Child physical aggression, Child verbal aggression, and Child negative act (nonaggressive).
- *Peer aggression*. Sum of Child physical aggression and Child verbal aggression.
- *Child noncompliance*. Sum of Says no/refuses to adult and Acts defiant to adult.

Qualitative Ratings

The Qualitative Ratings were designed to capture the quality of the child's caregiving experience. Each set of qualitative ratings is based on a complete 44-minute cycle, of which 25 minutes are designated for observing quality. Notes about quality are taken during the first 34 minutes of the observation. During the last 10 minutes of the observation, the observer focuses completely on observing quality and determining overall ratings based on the complete cycle (44 minutes). Observers rate quality items on a 4-point scale from 1, "not at all characteristic," to 4, "very characteristic." A more detailed description of each qualitative scale can be found in the corresponding Instrument Documents and Manuals (<http://secc.rti.org/>).

6 Month Qualitative Scales

- *Caregiver Notes*
 - Sensitivity/responsivity to distress
 - Sensitivity/responsivity to nondistress
 - Intrusiveness
 - Detachment/disengagement
 - Stimulation of development
 - Positive regard for the child
 - Negative regard for the child
 - Flatness of affect
- *Child Notes*

- Positive mood
- Negative mood
- Activity level
- Sociability
- Sustained attention

At the 6 month observation, the NICHD Early Child Care Research Network (1996, p. 278) created a composite of the Qualitative Ratings:

- *Positive caregiving ratings*. Sum of Sensitivity or Responsiveness to Non-distressed communication, positive regard, stimulation of cognitive development, detachment (reverse coded), and flat affect (reverse coded).
- The qualitative composite did not include ratings of intrusiveness or negative regard because extensive pilot observations indicated little variability in these rarely observed domains (NICHD Early Child Care Research Network, 1996, p. 282).

15 and 24 & 36 Month Qualitative Scales

All items are the same as those included in the 6 month measure with the addition of the following item:

- *Child Notes*
 - Positive engagement with caregiver

54 Month Qualitative Scales

- *Caregiver Ratings*
 - Sensitivity/responsivity
 - Intrusiveness/overcontrol
 - Detachment/disengagement
 - Stimulation of cognitive development
- *Child Ratings*
 - Self-reliance
 - Aggression/angry affect
 - Attention
 - Positive affect/mood
 - Activity
 - Social withdrawal from peers
- *Setting Ratings*
 - Chaos
 - Overcontrol
 - Positive Emotional Climate
 - Negative Emotional Climate

At the 54 month observation, the Qualitative Ratings generated the following composites (according to the NICHD Early Child Care Study Phase II Instrument Document):

- *Setting qualitative composite.* Sum of 4 settings ratings (Chaos + Overcontrol + Positive emotional climate + Negative emotional climate) after reverse coding Chaos, Overcontrol, and Negative emotional climate.
- *Caregiver qualitative composite.* Sum of 4 caregiver ratings (Sensitivity/responsiveness + Intrusiveness/overcontrol + Detachment/disengagement + Stimulation of cognitive development) after reverse coding Intrusiveness and Detachment.
- *Arrangement qualitative composite.* Sum of 4 settings ratings and 4 caregiver ratings after reverse coding Chaos, Overcontrol, Negative emotional climate, Intrusiveness, and Detachment.

Structural Variables

The observed structural variables capture environmental aspects of the caregiving environment.

Structural Variables (included in 6; 15; 24 and 36; and 54 month ORCE)

- *Ratio*
- *Group size*
- *Numbers of children*
- *Numbers of adults available*
- *Proportion of observation completed outdoors*
- *Amount of time caregiver is involved with child*
- *Age mix of the group*

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Trained observers conduct the observations. The procedure for observing includes two to four 44-minute observation cycles. Each cycle includes:

- 10 minutes of observation using the behavioral scales (using 30-second observe and record intervals)
- 2 minutes of note taking for qualitative ratings based on the preceding observation period
- 10 minutes of observing using the behavior scales followed by 2 minutes of note taking for qualitative ratings based on the preceding observation period
- 10 minutes of observation using the behavior scales
- 10 minutes of observation and note taking for the qualitative ratings that incorporate the preceding observation period and the current 10 minutes

Training Required: Approximately two days of training was provided on the ORCE in the NICHD Study of Early Child Care, followed by practice administering the instrument and tests of reliability to criterion coding (Bradley et al., 2003). "Data collectors were required to achieve at least 90% agreement with criterion coding to be certified. To maintain certification, data collectors were re-examined every 4 months using the same videotape procedure" (Bradley et al., 2003, p. 300).

Comment

The developers of the ORCE caution that unless a person has access to the NICHD training tapes, it would be difficult to use. There is no plan to release the tapes due to confidentiality issues. The developers note that without proper training reliability/validity of the ORCE in future use is not known.

Setting

Observations may be made in any non-maternal care environment.

Time Needed and Cost

Time: A cycle of recording consists of three 10-minute intervals of continuous recording, broken by 2-minute intervals for qualitative note-taking, followed by a 10-minute interval of observation focused on global qualities of behavior (that is, 44 minutes total for one observation cycle). Two to four such cycles of observation are collected at each assessment point (NICHD Study of Early Child Care Phase I Instrument Document, 2004, p.126-127).

Cost: The major cost involved is training observers to criterion. Administration of this pencil-and-paper instrument requires a timer.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

The research assistants who applied the ORCE throughout data collection during the Study of Early Child Care were certified at each data collection point through "gold standard" reliability coding. The gold standard reliability required that the research assistants' coding of taped caregiving situations, when compared to coding from master coders, achieved a passing score. Live reliabilities were also computed throughout each data collection period from the coding of two research assistants who coded the same caregiving situation. Two reliability estimates were computed from the gold standard and live codings, the Pearson Correlation and an estimate computed from the repeated measures ANOVA formulations provided in Winer (1971). The analysis variables that were reduced from ORCE cycles are divided into three categories; those from the behavioral items, those from the caregiver (adult) qualitative codes, and those from the child-centered qualitative codes.

The values reported in the following sections are the Median Pearson Correlations, as described above. The measure developers noted that sometimes different reliability estimates did not always indicate the same degree of reliability. For more specific information including all reliability estimates (Median Person Correlations, Median Winer Reliability, Pearson Correlation for "Live" data, Winer Reliability for "Live" data) for each variable, contact Bonnie Knoke (knoke@rti.org) at the Research Triangle Institute.

Please note that we have provided reliability information on composite variables that were created for the NICHD Study of Early Child Care. When available, we provided information on the construction of these composite variables in the Key Constructs section above. However, there are some reliability estimates reported for composite variables for which we did not have information on their construction. This information was not available in the materials that we had to review.

Behavioral Variables

6 Month Behavior Scales

Median Pearson Correlations by variable ranged from 0.41 to 0.99, with most estimates falling above 0.80. The reliability estimate for Stimulates Social Development and Negative Interaction were the lowest at 0.41 and 0.49, respectively.

15 Month Behavior Scales

Individual reliabilities for each variable are not reported for the 15 month data. Instead, reliability estimates are reported for 12 behavioral composites. The composites include Total Stimulation, No Stimulation, Response to Distress, Responsiveness, Negative Contact, Positive Contact, Rate of Physical Care, Child's Contact with Peers, Total Adult Attention, Total Adult Talk, and Group Interactions, and a total 15-month Behavioral Composite. All reliability estimates range from .64 to .93, with Positive Contact, Negative Contact, Response, and Peer Contact being the lowest.

24 Month Behavior Scales

Out of a total of forty-nine behavioral variables, all had acceptable levels of reliability with the exception of the following variables with low reliability and/or low frequency: Adult: Speaks Negatively to Child, Teaches Social Rule, Positive Physical Contact, Negative/Restrictive Actions, Mutual Pretend Play, Child: Prosocial Act, Negative (non-aggressive) Act, Verbal Aggression, Physical Aggression, Says No/Refuses Adult, Acts Defiant, Negative Behavior Toward Child. The composites with low levels of reliability were Adult: Proportion of Positive/Negative Behavior Toward Child, Child: Proportion of Compliance, Autonomy Proportion, Defiance Proportion, High Level of Peer Play, Proportion of Negative Peer Interaction, and Total Child Aggression.

36 Month Behavior Scales

Although the 24 and 36 month ORCE are the same, reliabilities were calculated at each time point. Median Pearson Correlations for individual variables ranged from 0.08 to 0.99. All of the variables had acceptable levels of reliability with the exception of the following variables: Adult: Speaks Positively to Children, Speaks Negatively to Children, Teaches Social Rule, Negative Restrict Actions, Child: Activity Alone, Activity Without Objects, Verbal Aggression, Says No/Refuses.

Additionally, 18 behavioral composites were created. The ORCE developers write that, "four variables have such low 'gold standard' reliability estimates along with

low Pearson correlations that ratification decisions for these variables at this time point should be cautiously made. The four composite variables are Negative Restricts Action + Speaks Negatively to Child, Negative Restricts Actions + Speaks Negatively to Child/Activity with Child or Adult, Complies with Adult/Gives Direction to Child, and Says No/Asks Questions of Children + Gives Directions to Child."

54 Month Behavior Scales

Median Pearson Correlations ranged from 0.34 to 0.97. Four variables had correlations of less than 0.60: Adult: Negative Management, Teaches Academic Skill, Facilitates Learning, and Child: Says No/Refuses to Adult. Five behavioral composites were created and range in reliability estimates from 0.20 (Child Noncompliance) to 0.95 (Peer Agonism). The developers note that the reliability estimates for Child Noncompliance were low because one of the two components was not observed in the data.

Qualitative Ratings

6 Month Qualitative Scales

- Adult: Median Pearson Correlations ranged from 0.55 to 0.94. All estimates were above 0.80 with the exception of Intrusiveness (0.55).
- Child: The reliabilities for Positive Mood, Negative Mood, and Positive Interaction with Peers fell between 0.69 and 0.89 indicating an acceptable level of reliability with these items. The reliabilities for Sociability and Sustained Attention were relatively low (0.45 and 0.51), and the reliability for Negative Interaction with Peers was considered low by the developers at 0.49. They note that this estimate was particularly surprising at 6 months.

15 Month Qualitative Scales

- Adult: Median Pearson Correlations ranged from 0.20 to 0.85. The estimates for Negative Regard for Child (0.20) and Intrusiveness into Child's Activities (0.47) were relatively low.
- Child: The child qualitative variables had lower reliabilities than the adult reliabilities. All estimates were between 0.34 and 0.77. Only Positive Engagement with Caregiver (0.77) and Sociability (0.77) had acceptable levels of reliability.

24 Month Qualitative Scales

- Adult: Negative regard, Intrusion, and Flatness of Affect all showed poor reliability, with the remaining variables and two composite variables showing adequate reliability.
- Child: All estimates were between 0.47 and 0.76. Only Positive Engagement with Caregiver and Positive Mood had acceptable levels of reliability.

36 Month Qualitative Scales

- Adult: A composite variable including the eight individual variables was created and showed adequate reliability (0.80). Three of the individual variables, flatness of affect, fostering exploration, and intrusion showed relatively low reliability (0.32 – 0.57).
- Child: The Pearson Correlations ranged from 0.57 – 0.93. The estimate for Child's Level of Negative Mood was the lowest (0.57). Only two variables showed acceptable levels of reliability: Child's Activity Level and Child's Level of Positive Mood.

54 Month Qualitative Scales

- Adult: The Pearson Correlations ranged from 0.62 (Caregiver Detachment/Disengagement) to 0.76 (Caregiver Sensitivity/Responsivity). The caregiver composite had a reliability estimate of 0.73.
- Child: The Pearson Correlations ranged from low (0.20, Child Social Withdrawal) to high (0.83, Child Aggression/Angry Affect). The setting composite had a reliability of 0.75 and the arrangement composite 0.77.

Validity Information

Construct Validity

ORCE measures of child care "quality" were related to expectable structural variation such as level of caregiver education and adult-child ratio (see references below).

NICHD Early Child Care Research Network. (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly*, 11, 269-306.

NICHD Early Child Care Research Network. (2000). Characteristics and quality of child care for toddlers and preschoolers. *Applied Developmental Science*, 4, 116-135.

NICHD Early Child Care Research Network. (2002). Child-care structure → process → outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science*, 13, 199-206.

Concurrent Validity

ORCE measures of child care "quality" were related to expectable child measures such as cognitive performance, language, and social functioning (see references below).

NICHD Early Child Care Research Network. (2000). The relation of child care to cognitive and language development. *Child Development*, 71, 960-980.

NICHD Early Child Care Research Network. (2002). Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Educational Research Journal*, 39, 133-164.

NICHD Early Child Care Research Network. (2003). Does quality of child care affect child outcomes at age 4½? *Developmental Psychology*, 39, 451-469.

Predictive Validity

ORCE measures of child care "quality" were related to later child outcomes in the areas of cognitive performance, language, and social functioning (see references below).

NICHD Early Child Care Research Network. (2003). Social functioning in first grade: Associations with earlier home and child care predictors and with current classroom experiences. *Child Development*, 74, 1639-1662.

NICHD Early Child Care Research Network. (2002). Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Educational Research Journal*, 39, 133-164.

NICHD Early Child Care Research Network. (2003). Does quality of child care affect child outcomes at age 4½? *Developmental Psychology*, 39, 451-469.

Comment

The Modified ORCE (M-ORCE) was created using items from the 24/36-month, and 54-month versions of the Observational Record of the Caregiving Environment (ORCE) to "reflect caregiver and child behaviors that were appropriate for scoring across a wide age range. For example, the 24- and 36-month ORCE differentiates between caregiver sensitivity to distress and sensitivity to non-distress, while the 54-month ORCE does not. To be more inclusive across a wide age range, the M-ORCE uses the more general 54-month ORCE sensitivity ratings. In contrast, the 24-/36-month ORCE has codes for the care provider speaking positively and negatively to the child, while this code is not available on the 54-month ORCE. The M-ORCE contains the more general positive versus negative talk items. In addition, several new codes and ratings were developed to reflect the quality of the child's functioning at child care. . . Definitions of these codes include the extent to which the child was integrated into positive social activities or was only on the fringe of social groups, the care providers use of strategies that support or impede the development of a sense of community among the children, and measures of the quality of peer relations. These codes and ratings were designed to capture more qualitative aspects of the child's actions at child care" (Kryzer et al. 2007, p. 454).

The manual for the M-ORCE is available, free of cost, from the developers. However, training on this measure is not available. For more information, please contact Megan Gunner at gunnar@umn.edu.

References and Additional Resources

- Abbot – Shim, M., & Sibley, A. (1987). *Assessment Profile for Early Childhood Programs*. Atlanta, GA: Quality Assist.
- Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology, 10*, 541-552.
- Bradley, R., Caldwell, B. & Corwyn, R. (2003). The Child Care HOME Inventories: Assessing the Quality of Child Care Homes. *Early Childhood Research Quarterly, 18*(3), 294-309.
- Harms, T., & Clifford, R. M. (1980). *Early Childhood Environment Rating Scale*. New York, NY: Teachers College Press.
- Harms, T., & Clifford, R. M. (1989). *Family Day Care Rating Scale*. New York, NY: Teachers College Press.
- Kryzer, E. M., Kovan, N., Phillips, D. A., Domagall, L. A., & Gunnar, M. R. (2007). Toddlers' and preschoolers' experience in family day care: Age differences and behavioral correlates. *Early Childhood Research Quarterly, 22*(4), 451-466.
- NICHD Study of Early Child Care and Youth Development Phase I Instrument Document (<http://secc.rti.org/instdoc.doc>). (2004). Observational Record of the Caregiving Environment (ORCE): Behavior Scales, Qualitative Scales, and Observed Structural Variables for 6, 15, 24, & 36 months.
- NICHD Study of Early Child Care and Youth Development Phase II Instrument Document (<http://secc.rti.org/Phase2InstrumentDoc.pdf>). Observational Record of the Caregiving Environment (ORCE): Behavior Scales, Qualitative Scales, and Observed Structural Variables for 54 months.
- NICHD Early Child Care Research Network (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly, 11*, 269-306.
- NICHD Early Child Care Research Network. (2001). Nonmaternal care and family factors in early development: An overview of the NICHD Study of Early Child Care. *Journal of Applied Developmental Psychology, 22*, 457-492.

Program Administration Scale (PAS)

I. Background Information

Author/Source

Source: Talan, T. N. & Bloom, P. J. (2004). *Program Administration Scale: Measuring Leadership and Management in Early Childhood Programs*. New York, NY: Teachers College Press.

Publisher: Teachers College Press
1234 Amsterdam Avenue
New York, NY 10027
www.teacherscollegepress.com

Purpose of Measure

As described by the authors:

"The Program Administration Scale (PAS) was designed to serve as a reliable and easy-to-administer tool for measuring the overall quality of administrative practices of early care and education programs and as a useful guide to improve programs" (Talan & Bloom, 2004, p. 1).

"The PAS was constructed to complement the widely used observation-based classroom environment rating scales designed by Harms, Clifford, and Cryer...If used together these instruments provide a focused look at best practices at the classroom level and the broad view of program quality from an organizational perspective" (Talan & Bloom, 2004, p. 1).

Population Measure Developed With

The PAS was "designed for early childhood program administrators, researchers, monitoring personnel, and quality enhancement facilitators" (Talan & Bloom, 2004, p. 1). Reliability and validity was established with a sample from early care and education programs in Illinois.

Age Range/Setting Intended For

The PAS was developed for use in center-based or public school-based early care and education programs.

Ways in which Measure Addresses Diversity

The PAS includes a subscale examining family partnerships. The subscale includes items and indicators that assess communication with families about values, beliefs and cultural practices.

Key Constructs & Scoring of Measure

The *Program Administration Scale* measures both leadership and management functions of early childhood administration. Leadership functions include clarifying

values, articulating a vision, setting goals, and charting a course of action to achieve those goals. Management functions relate to orchestrating tasks and developing systems to carry out the organizational mission.

The PAS includes 25 items that are clustered in 10 subscales. The subscales and items are as follows:

- *Human Resources Development* (3 items)
 - Staff orientation
 - Supervision and performance appraisal
 - Staff development
- *Personnel Cost and Allocation* (3 items)
 - Compensation
 - Benefits
 - Staffing patterns and scheduling
- *Center Operations* (3 items)
 - Facilities management
 - Risk management
 - Internal communications
- *Child Assessment* (2 items)
 - Screening and identification of special needs
 - Assessment in support of learning
- *Fiscal Management* (2 items)
 - Budget planning
 - Accounting practices
- *Family Partnerships* (2 items)
 - Family communications
 - Family support and involvement
- *Program Planning and Evaluation* (2 items)
 - Program evaluation
 - Strategic planning
- *Marketing and Public Relations* (2 items)
 - External communications
 - Community outreach
- *Technology* (2 items)
 - Technological resources
 - Use of technology
- *Staff Qualifications* (4 item)
 - Administrator
 - Lead Teacher
 - Teacher
 - Apprentice Teacher/Aide

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Director, assistant director, or program administrator, and trained independent assessors, such as researchers, consultants and program evaluators.

Training Required: The McCormick Center for Early Childhood Leadership provides four types of training experiences which they describe on their website (<http://cecl.nl.edu>). The trainings are described as follows:

- Widening the Lens – When the goal is expanding awareness of the importance of assessing quality from a total organizational perspective. This 1-2 hour overview training is designed for center directors, managers, technical assistance specialists, college instructors, researchers, and policymakers.
- Leading the Way to Quality – When the goal is program self-improvement. This 1-2 day training is designed for center directors and managers and is delivered in an interactive workshop format.
- Supporting Directors as the Gatekeepers to Quality – When the goal is quality facilitation. This 2-3 day training is designed for technical assistance specialists, mentors, and consultants who are involved in quality enhancement initiatives and is delivered in an interactive workshop format.
- PAS Assessor Reliability Training – When the goal is research, evaluation, or monitoring quality. This 4 day intensive training seminar is designed for researchers, program evaluators, technical assistance specialists, and consultants.

All trainings can be customized to meet needs and delivered on-site. To schedule a training, contact Jill Bella, Director of Special Projects at (800) 443-5522, ext. 5059 or jill.bella@nl.edu.

Setting

For formal assessments, interviews are set-up with the program administrator at the site, which enables access to required documents for the visit.

Time Needed and Cost

Time: For formal assessments, the authors recommend a time frame of two hours for an interview with the on-site administrator and an additional 2 – 4 hours for a review of required documents.

Cost: The PAS is \$19.95 and can be purchased from New Horizons or Teachers College Press. A new book is needed each time the PAS is administered.

For more information on the cost of training, please visit The McCormick Center for Early Childhood Leadership website (<http://cecl.nl.edu>).

III. Functioning of Measure

Reliability Information

"A reliability and validity study of the PAS was conducted in 2003 involving 67 center-based early childhood programs. Data generated from the reliability and validity study were used to make revisions in the wording of different indicators, delete redundant items, and streamline the data-collection protocol" (Talan & Bloom, 2004, p. 69).

A pool of 176 programs in Illinois was developed representing urban, suburban, and rural geographic regions as well as programs varying by accreditation status and size of the center (small, medium, and large). From the pool, 124 centers were randomly contacted and invited to participate in the reliability and validity study. Slightly more than half (67) agreed to participate. The participating centers were equally split between accredited (48%) and not accredited (52%). Two-thirds of the participating programs were nonprofit. It should be noted that the 67 programs participating in the pilot did not receive a copy of the instrument prior to the administration of the scale by a trained assessor. It is anticipated that as the PAS is used broadly, the percentage of programs scoring a level 1 on items will decrease as on-site administrators prepare the documentation needed to verify each indicator.

Inter-rater Reliability

"Inter-rater reliability was determined during training on the use of the instrument with eight early childhood assessors. Using a videotape of the entire interview protocol, assessors were rated on how often they matched the PAS anchor's scores within 1 point for each item. Individual assessor's inter-rater reliability scores ranged from 81% to 95% agreement on the 25 items. Overall inter-rater reliability was 90% for the eight assessors used in the PAS pilot" (Talan & Bloom, 2004, p. 72).

Internal Consistency

Internal consistency was determined through computation of Cronbach's Alpha coefficient on the Total PAS scores from the 67 sites in the reliability and validity study. Coefficient alpha for the Total PAS was .85 (Talan & Bloom, 2004). A more recent study by Lower and Cassidy (2007) found a coefficient alpha of .88 for the first nine subscales of the PAS. The final subscale measuring Staff Qualifications was not used in the analyses because information was not consistently reported for all classroom teachers, which is needed to accurately complete this subscale.

Validity Information

Content Validity

Content validity was established by a panel of 10 early childhood experts who evaluated each indicator, item, and subscale on the PAS to ensure the key leadership and management practices of center-based early childhood programs were included (Talan & Bloom, 2004, p.70). It was also reviewed informally by 10 other early childhood administrators, consultants and trainers.

Construct Validity

The 10 subscales were correlated to determine the extent to which they measured distinct, though somewhat related, aspects of early childhood administration. Subscale intercorrelations ranged from .09 to .63, with a median value of .33, confirming that the subscales, for the most part, measure distinct characteristics of organizational administration.

Criterion Validity

The authors used accreditation status as a proxy for program quality and compared Total PAS scores between programs accredited by the National Association for the Education of Young Children and those not currently accredited. Accredited programs had higher scores on the PAS ($M = 92.12$, $S.D. = 19.43$) than not-accredited programs ($M = 72.06$, $S.D. = 20.83$) (ANOVA, $F = 16.59$, $p < .01$).

Concurrent Validity

Concurrent validity was determined through a correlational analysis with two other instruments that measure early childhood organizational effectiveness: the Early Childhood Work Environment Survey (ECWS; Bloom, 1996) and the Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms, Clifford, & Cryer, 1998). Lower and Cassidy (2007) found a statistically significant moderate correlation ($r(54) = .291$, $p = .031$) between the PAS and global classroom quality measured by the ECERS-R. Additionally the authors found a statistically significant positive correlation ($r(88) = .287$, $p = .006$) between the PAS and the Parents and Staff Subscale of the ECERS-R combined with ratings from the Infant/Toddler Environment Rating Scale-Revised (ITERS-R; Harms, Cryer, & Clifford, 2003). A positive correlation ($r(25) = .331$, $p = .098$) was also found between the PAS and the Organizational Climate scale of the ECERS-R.

Brooks-Gunn et al. (2008) also found significant correlations between the PAS and the ECERS-R with a Pearson r of .52 with a significance level of $p < .01$. Additionally, the authors found a significant relationship between the PAS and the Program Structure Subscale of the ECERS-R; however, no relationship was found between the PAS and the Parents and Families Subscale of the ECERS-R. Additional analyses were performed to determine the relationship between the PAS and a combination of the Program Structure and Parents and Families Subscales of the ECERS-R. Results showed significant positive correlations between the PAS and the combined subscale with a Pearson r of .41 and a significance level of $p < .05$. Talan and Bloom (2004, p. 73) also found moderate correlations with both the ECERS-R and ECWES, which indicates that the PAS measures constructs that are related to but are not the same as characteristics of organizational quality.

References and Additional Resources

Brooks-Gunn, J., Kagan, S. L., Tarrant, K., Cortazar, A., Johnson, A., Philipsen, N., et al. (2008). *New York City early care and education unified performance measurement system*. New York, NY: The New York City Administration for Children's Services.

- Bloom, P. J. (1996). *Improving the quality of work life in the early childhood setting: Resource guide and technical manual for the Early Childhood Work Environment Survey*. Wheeling, IL: McCormick Center for Early Childhood Leadership, National-Louis University.
- Harms, T., Cryer, D., Clifford, R. M. (2003). *Infant/Toddler Environment Rating Scale: Revised Edition*. New York, NY: Teachers College Press.
- Harms, T., Clifford, R., and Cryer, D. (1998). *Early Childhood Environment Rating Scale – Revised*. New York: Teachers College Press.
- Lower, J. K., & Cassidy, D. J. (2007). Child care work environments: The relationship with learning environments. *Journal of Research in Childhood Education*, 22(2), 189-204.
- Talan, T. N. & Bloom, P. J. (2004). *Program Administration Scale: Measuring Leadership and Management in Early Childhood Programs*. New York, NY: Teachers College Press.

Program for Infant/Toddler Care Program Assessment Rating Scale (PITC PARS)

I. Background Information

Author/Source

Source: Kriener-Althen, K. & Mangione, P. (in preparation). *PITC PARS Technical Manual*. San Francisco, CA: WestEd.

Kriener-Althen, K., Niggle-Hollis, M., & Mangione, P. (in preparation). *PITC PARS User's Guide*. San Francisco, CA: WestEd.

Mangione, P. (in press). *Program for Infant/Toddler Care Program Assessment Rating Scale (PITC PARS)*. San Francisco, CA: WestEd.

Mangione, P. (in press). *Program for Infant/Toddler Care Family Child Care Program Assessment Rating Scale (PITC FCC PARS)*. San Francisco, CA: WestEd.

Mangione, P., Kriener-Althen, K., Niggle, M. P., & Welsh, K. (2006). *Program Quality Through the PITC Lens: Assessing Relationship-Based Care in Infant/Toddler Early Care and Education Programs*. Presentation at the 15th National Institute for Early Childhood Professional Development. San Antonio, TX.

WestEd. (2006). A new yardstick for assessing early childhood care. *R & D Alert*, 8(1), 18-21. San Francisco, CA: WestEd.

Publisher: WestEd
730 Harrison Street
San Francisco, CA 94107-1242
Phone: 415-565-3000
Website: www.WestEd.org

Purpose of Measure

As described by the authors:

The Program for Infant/Toddler Care Program Assessment Rating Scale (PITC PARS) is an observational instrument designed to assess the quality of early care and education settings for infants and toddlers. The PITC PARS measures the extent to which caregiving practices, the care environment, program policies and administrative structures promote responsive, relationship-based care for infants and

toddlers. The PITC PARS utilizes a positive orientation to assessing various aspects of program quality.

The PITC PARS is based on constructs developed for the Program for Infant/Toddler Care, a comprehensive multimedia training system for infant/toddler care teachers. The PITC constructs were developed through a multi-step process that included a literature review of early development and child care quality and consultation with an advisory panel of national experts in the fields of infant development and early care and education. The literature on the assessment of child care quality was reviewed and extensive piloting was conducted to support the development of the PITC PARS. In addition to the PITC constructs, the Observational Record of Classroom Observation Environments (ORCE, ECRN, NICHD, 1996) and recommended practices of the American Academy of Pediatrics provided the foundation for the development of the PITC PARS.

The instrument was originally designed to assess the implementation of the PITC approach to infant/toddler care, but may also be generally used for program development and evaluation assessment.

Population Measure Developed With

The PITC PARS has been used to assess the implementation of the PITC training and technical assistance in five evaluation studies in California. Psychometric information for the PITC PARS was based on data collected with:

- 1,071 infant/toddler care teachers (858 in center-based programs; 213 in family child care programs)
- 735 classrooms (546 center-based; 189 family child care)

The PITC PARS has also been used in evaluation studies of the PITC in Iowa, South Dakota, and Oklahoma to assess the effectiveness of infant/toddler training and technical assistance. In addition, a subset of PARS items has been included in the study of the implementation of infant/toddler care quality standards in Indiana.

Age Range/Setting Intended For

The PITC PARS was created for use in infant/toddler care programs, i.e., those that serve children age birth to three outside the home. There is a version of the instrument available for use in center-based programs (PITC PARS) and a version available for use in family child care (PITC FCC PARS). The two versions of the instruments were designed to allow comparison of assessment data between them.

Ways in which Measure Addresses Diversity

Items that assess cultural responsiveness and representative staffing are included in Subscale II of the PITC PARS.

The PITC PARS has been validated in diverse infant/toddler care settings representative of state-subsidized programs. Settings with diverse groups of children

and diverse infant care teachers have been assessed. The diversity includes both linguistic and cultural dimensions as well as urban, suburban, and rural locations.

Key Constructs & Scoring of Measure

The user's guide describes the PITC PARS as —a scale with 5 sub-scales, 27 items, and 108 sub-items. This structure suggests a superficial understanding of PITC, which is a holistic philosophy. However, the intent of this structure is to measure distinct aspects of PITC. The specificity at the sub-item level facilitates reliable and valid measurement of each item. Sub-item scores are aggregated to create item scores, which are aggregated to create sub-scale and scale scores” (Kriener-Althen, Niggle-Hollis, & Mangione, in preparation, p. ii).

Each sub-scale is comprised of four or more items. An item measures a specific aspect of quality within each sub-scale. There are four sub-items within each item. Each sub-item measures a specific aspect of quality defined under each item. Together, the ratings on the four sub-items provide an overall assessment of the extent to which the item has been implemented by infant/toddler care teachers and/or programs.

The five subscales are as follows:

- I. *Quality of Caregivers' Interaction with Infants*— Subscale I consists of 7 items (28 sub-items). It assesses the responsiveness of individual care teachers' interactions with the infants and toddlers in their care.
- II. *Family Partnerships, Cultural Responsiveness, and Inclusion of Children with Disabilities and Other Special Needs* – Subscale II consists of 5 items (20 sub-items). It assesses the extent to which the individual needs and preferences of infants/toddlers and their families are reflected in caregiving practices, the classroom environment, and program policies.
- III. *Relationship-Based Care* – Subscale III consists of 4 items (16 sub-items). It assesses the extent to which caregiving practices and program procedures meet infants' and toddlers' individual needs by establishing predictable and supportive relationships with 1 or 2 care teachers.
- IV. *Physical Environment* – Subscale IV consists of 7 items (28 sub-items). It assesses the extent to which a program provides indoor and outdoor environments that support infants' and toddlers' developmental needs for exploration, movement, and appropriate play materials.
- V. *Routines and Record Keeping* – Subscale V consists of 4 items (16 sub-items). It assesses the extent to which caregiving routines and program procedures promote infants' and toddlers' safety and health.

Each sub-item is assessed as "met" or "not met." Guidelines for determining "met" or "not met" are provided for each sub-item in the *PITC PARS User's Guide*. "Met" items are assigned a value of "1" and summed to produce item scores. Item scores are averaged to produce sub-scale scores. Each PITC PARS' sub-scale is scored on a scale ranging from 0 to 4 (see Table 1).

TABLE 1

Interpretations of PARS Subscale Ratings

Minimal Score by Level	INTERPRETATION
0 ³	Inadequate
1.8 ⁴	Minimal
2.8 ⁵	Good
4 ⁶	Excellent

Comments

The PITC FCC PARS includes adaptations to some PITC PARS sub-items appropriately applicable to family child care settings. Interpretation of sub-items for center-based and home-based settings are specified in the *PITC PARS User's Guide*.

The authors recommend developing individual ratings on Subscale I for each infant/toddler care teacher in the room, to provide the most precise measurement of the interactions. Subscale I was previously used to develop one rating for Subscale I to provide an impression of the quality of all interactions between teachers and children in the room during the observation. However, over time, observers provided consistent feedback of the challenges of reliably rating classrooms where one care teacher was responsive and another was less responsive, or classrooms where assistant teachers or others were in the classroom for short periods of time. Rating a classroom rather than individual teachers also proved challenging for obtaining inter-rater reliability and for distilling the effects of training over time. For these reasons, it is recommended that Subscale I be completed for individual care teachers, to capture each teacher's strengths when interacting with children.

According to the user's guide, the scale was originally "constructed using sentences with a 5-point likert scale;" however, as PITC trainers and others looked at the items in the scale, they felt that important pieces were missing. Descriptive text was added to each item, which often meant that observers' ratings were based on fairly lengthy paragraphs. This structure made achieving inter-rater reliability difficult. Consequently, the structure was then revised to its current form – the content in each of the item paragraphs was segmented into four distinct sub-items. Each sub-item is rated as either "Met" or "Not Met," with the sub-items functioning as a checklist within each item (Kriener-Althen, Niggle-Hollis, & Mangione, in preparation).

³ Programs at this level are meeting 45% or less of the total number of items.

⁴ Programs at this level are meeting 45-70% of the total number of items.

⁵ Programs at this level are meeting 70-99% of the total number of items.

⁶ Programs at this level are meeting 100% of the total number of items.

II. Administration of Measure

Who Administers Measure/Training Required

The PITC PARS has most frequently been used by program evaluators and researchers to determine the extent to which PITC's essential policies are implemented within classrooms and programs, and the extent to which a program's quality improves after participating in PITC onsite training and technical assistance plans. In addition, it has been utilized more generally to assess program quality in evaluation studies. The validity data reported later in this summary indicate that the PITC PARS correlates with commonly used measures of program quality. In studies beyond the PITC training and implementation context, the PITC PARS provides evaluators and researchers with a broad set of items that focus on the quality of infant care teachers' interactions with children.

Test Administration: It is recommended that trained, independent assessors administer the PITC PARS. According to the user's guide, the PITC PARS assesses the quality of relationships between infants and toddlers and their earliest caregivers "through observation, interview and review of the program's written materials. Subscale I 'Quality of Caregivers' Interaction with Infants' is solely rated through observation, which provides unique challenges for new users to the instrument. Rating with the PARS requires the ability to document interactions over the time span of a few hours. This documentation is used later, during rating, to create, as objectively as possible, an overall impression of the caregiver. Specifically, observers document specific teacher-child interactions, including dialogue with as many direct quotes as possible, and children's reactions. This level of documentation facilitates the observer's ability to form an overall impression of the care when rating. As with all observational assessments, the PITC PARS requires a certain level of judgment. The *PITC PARS User's Guide* is provided to assist users in developing this capacity to judge and to consistently apply principles with objective judgment" (Kriener-Althen, Niggle-Hollis, & Mangione, in preparation, p. ii).

Training Required: It is recommended that users receive training from PITC PARS data anchors at WestEd, Center for Child and Family Studies Evaluation Team. PARS training can be customized to support individualized needs, but generally consists of the four components described below. The number of estimated days to achieve reliability applies separately to each version of the PARS, so, for organizations interested in both versions (for family child care and center-based infant/toddler care), the number of days would need to be doubled.

- **Introduction/Orientation to the PITC PARS** – This component usually includes at least one day of classroom-style training that can accommodate 30 or more participants; and 2 or more days of initial field training with a limited number of participants. The purpose of this component is to provide an overall introduction and orientation to the use of the instrument. Content covered includes the philosophy underlying the instrument, structural components, mechanisms for use, strategies for administration, and scoring.

- **Practicing with the PITC PARS** – Individuals who will be using the PARS are encouraged to gain experience with the instrument on their own and to discuss their experiences with their colleagues, in preparation for interrater reliability training with PITC PARS anchors.
- **Inter-rater Reliability Training** – This component includes at least one day of classroom-style training day to discuss questions from individual practice and at least three days of inter-rater reliability observations in the field with a PITC PARS anchor and a limited number of participants. A person is considered reliable when a minimum standard of exact agreement for 80% or more sub-items across three successive observations is achieved.
 - Standards for achieving inter-rater reliability are that trainees must conduct three consecutive observations for which they have exact agreement with the WestEd PARS anchor for at least 80% of the sub-items. For example, if the inter-rater reliability is below 80% on the third observation, the timeline is reset, and training continues until the trainee achieves 80% reliability on three consecutive observations. These standards for inter-rater reliability are applied independently to the PITC PARS and PITC FCC PARS instruments.
 - To achieve inter-rater reliability, the WestEd PARS anchor and trainee assess the same teacher(s) and classroom and complete ratings independently of each other. The anchor and trainee then complete an inter-rater reliability score sheet to establish how many sub-item ratings were exactly matched. They then begin a reliability discussion with priority given to those sub-items where exact match was not achieved. The trainee and the WestEd PARS anchor each present evidence compiled during the observation to support their ratings. During this time, trainees gain familiarity with the tool and insight into the procedures the WestEd PARS anchor uses to determine a rating. This discussion is often lengthy, but essential to reaching consensus.
 - Upon achieving reliability, the trainee is eligible to train local assessors for their own evaluation and research projects. Suggestions for maintaining high inter-rater reliability with their assessors are to hold weekly assessor meetings during periods of active data collection, consult the *PITC PARS User's Guide* for determination of sub-item ratings, and conduct regular inter-rater reliability checks as frequently as after every 10 to 12 observations to maintain consistency among all assessors.
- **Follow-Up Assistance** – Follow-up assistance is available for organizations needing additional sessions to achieve inter-rater reliability and/or consultation to train local evaluation/research staff.

Setting

Observations, interviews, and review of program materials are conducted in infant/toddler care programs in center-based and/or family child care programs.

Time Needed and Cost

Time: The authors recommend conducting observations in the care environment for a minimum of 3 hours, followed by an interview with a program administrator, and review of written program materials. The assessment should be scheduled to include opportunities to observe the following activities: interactions with parents when children are dropped off or picked up from care, routines, meals, and indoor and outdoor play. Individual ratings on Subscale I can be made for up to three care teachers from the same classroom during one observation period. However, it is recommended that only one classroom be assessed per observation period, and that ratings are completed for one classroom before the next classroom is observed.

Cost: Costs are dependent upon options chosen, number of anchors trained, travel costs, and whether training on only one or both versions of the instrument (PARS and FCC PARS) is desired. Contact WestEd, Center for Child and Family Studies Evaluation Team for information about costs of training and materials.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

In practice, inter-rater reliability scores with an exact match at the sub-item level ranged from 79% (Subscale III) to 86% (Subscale II; Mangione et al., 2006).

Internal Consistency

Cronbach's alpha reliability statistics were computed for subscales at the sub-item level. Internal reliability scores were high overall: 0.90 for Subscale I, 0.76 for Subscale II, 0.74 for Subscale III, 0.80 for Subscale IV, and 0.70 for Subscale V.

Confirmatory factor analyses were performed separately for the PITC PARS and PITC FCC PARS. The following three factors were independently identified in each analyses: (1) interactions/relationships with infants; (2) policies with families, culture, inclusion, and primary caregiving; (3) the organization and structure of group care through environment and routines.

Validity Information

Concurrent Validity

The PITC PARS has been used in a complementary way with the Environmental Rating Scales (ERS) and the Arnett Scale of Caregiving Behavior. Overall correlations between the PITC PARS and the ERS have been high, ranging from 0.81 on the FDCRS to 0.88 on the ECERS-R. Correlations between the PITC PARS Subscale I and the Arnett Scale of Caregiving Behavior have been moderately high, ranging from 0.60 on the Arnett Warmth subscale to -0.70 on the Arnett Criticalness subscale (Mangione, et al, 2006).

Predictive Validity

The PITC PARS was used in a pre-post analysis of infant/toddler care teachers from center-based and family child care settings who participated in PITC onsite training and technical assistance between 2000 and 2002. Statistically significant improvements were documented in the overall quality of three samples of programs (two center-based samples and one family child care sample) completing training and technical assistance plans. Overall improvements were identified; however, the quality of the care teachers' interactions with infants and toddlers (Subscale I) demonstrated the most consistent positive change (Mangione, 2003).

The PITC PARS was used in a repeated measures analysis of infant/toddler care teachers from center-based and family child care settings who participated in PITC onsite training and technical assistance between 2004 and 2007. Results identified statistically significant positive linear relationships between participation in PITC *Partners for Quality Training Plans* and improved quality of care in the areas of relationship-based care (Subscale III) and the physical environment (Subscale IV; Kriener-Althen & Mangione, 2007).

Content Validity

The PITC PARS is based on constructs developed for the Program for Infant/Toddler Care, a comprehensive multimedia training system for infant/toddler care teachers. The PITC constructs were developed through a multi-step process that included literature review of early development and child care quality and consultation with an advisory panel of national experts in the fields of infant development and early care and education (Lally & Mangione, 2008; Mangione, 1990). The literature on the assessment of child care quality was reviewed, and extensive piloting was conducted to support the development of the PITC PARS (Kriener-Althen & Mangione, in preparation).

References and Additional Resources

- Kriener-Althen, K. & Mangione, P. (in preparation). *PITC PARS Technical Manual*. San Francisco, CA: WestEd.
- Kriener-Althen, K. & Mangione, P. (2007). Monitoring PITC Partners for Quality Training and Technical Assistance. Excerpt from PITC Annual Report. Sausalito, CA: WestEd Center for Child and Family Studies.
- Kriener-Althen, K., Niggle-Hollis, M., & Mangione, P. (in preparation). *PITC PARS User's Guide*. San Francisco, CA: WestEd.
- Lally, J. R., & Mangione, P. L. (2008). The Program for Infant Toddler Care. In J. P. Roopnarine, & J. E. Johnson, (Eds), *Approaches to early childhood education, 5th Edition*. Englewood Cliffs, NJ: Prentice Hall.
- Mangione, P. (in press). *Program for Infant/Toddler Care Program Assessment Rating Scale (PITC PARS)*. San Francisco, CA: WestEd.

- Mangione, P. (in press). *Program for Infant/Toddler Care Family Child Care Program Assessment Rating Scale (PITC FCC PARS)*. San Francisco, CA: WestEd.
- Mangione, P. (2003). *Impact of PITC Training on Quality of Infant/Toddler Care, Evaluation Report*. Sausalito, CA: WestEd Center for Child and Family Studies.
- Mangione, P.L. (1990). A comprehensive approach to using video for training infant and toddler caregivers. *Infants and Young Children, 3* (1), 61-68.
- Mangione, P., Kriener-Althen, K., Niggle, M. P., & Welsh, K. (2006). *Program Quality Through the PITC Lens: Assessing Relationship-Based Care in Infant/Toddler Early Care and Education Programs*. (Presentation). 15th National Institute for Early Childhood Professional Development. San Antonio, TX.
- WestEd. (2006). A new yardstick for assessing early childhood care. *R & D Alert, 8*(1), 18-21. San Francisco, CA: WestEd.

The Preschool Classroom Implementation Rating Scale (PCI)

I. Background

Author/Source

Source: Frede, E. C. & Miller, A. K. (1990). *Preschool Classroom Implementation Rating Instrument – High/Scope Manual*.

Publisher: This measure is currently unpublished, but is available from the first author at efrede@tcnj.edu.

Purpose of Measure

The Preschool Classroom Implementation Rating Scale (PCI) was originally developed to measure treatment fidelity of the High/Scope curriculum. A shortened, more general version was later developed. "Embedded within the PCI is a subscale that measures general quality factors for a cognitive-developmental classroom. This subscale forms the basis for the PCI-CD (CD for Cognitive/Development) which adheres to the constructive philosophy of the original instrument but deletes those items which are specific to the High/Scope approach" (Frede & Miller, 1990, p. 1).

"The PCI is a checklist of adult behaviors or environmental factors determined by the adult. The behaviors included in the instrument are all deemed to be indicators of quality in a cognitive-developmental or constructivist classroom" (Frede & Miller, 1990, p. 1).

"The PCI differs from some other observation instruments in that it does not measure micro-level interactions, nor does it inventory teaching techniques which should be seen in classrooms using other approaches such as direct instruction. The PCI by itself does not provide information on aspects of quality other than teacher behavior or the learning environment" (Frede & Miller, 1990, p. 1).

Population Measure Developed With

The PCI was developed based on pre-school classrooms including children with disabilities and mixed family incomes.

Age Range/Setting Intended For

The PCI was developed for use in programs for children ages three through six. "It is appropriate for observing programs in pre-school or kindergarten classrooms in public or private schools, day care centers, Head Start, or church programs" (Frede & Miller, 1990, p. 2).

Ways in Which Measure Addresses Diversity

The instrument looks at specific teaching strategies but not content, so while strategies that help children develop social competence are included, specific anti-bias or multicultural education strategies are not included.

Key Constructs & Scoring of Measure

The PCI has 52 items in 12 subscales. Each item is given a rating of not observed, not evident, evident, or optimal. The subscales and items are as follows:

- *Room Arrangements* (8 items)
 - Activity areas are clearly defined
 - Traffic flow is not impeded by boundaries
 - Materials are logically arranged
 - Materials are labeled
 - Materials are easily accessible to children
 - There is sufficient amount of unstructured material in each area
 - Real tools and household equipment are available
 - A wide variety of books that are age-appropriate and always accessible in an inviting location in the room
- *Daily Routine* (4 items)
 - Time periods have specific names which children are helped to learn
 - Routine is consistent from day to day
 - Adults help children make transitions from one part of routine to another
 - The planning time, work time, recall time sequence is not interrupted
- *Planning Time* (3 items)
 - Adults meet with the same group of children daily
 - The planning process is made interesting and stimulating to the children
 - Individual children plan according to their ability
- *Work Time/Free Play* (2 items)
 - Work time is 45 minutes long
 - Children are involved in child-initiated activities and are free to move from one activity to another
- *Clean-Up Time* (2 items)
 - Adults use appropriate strategies to encourage clean-up
 - Adults take advantage of opportunities for incidental teaching of the key experiences
- *Recall Time* (2 items)
 - Adults have recall with the same group of children with which they are planned
 - Adults use a variety of strategies to make recall time interesting to children in their groups
- *Small Group Time* (5 items)
 - Adults have materials ready for small group time
 - Every child has his own materials at small group time
 - Small group activities allow each child to make choices
 - Small group activities have a key experience focus, but children respond according to their own abilities
 - Each small group is well-organized with a beginning, middle, and end

- *Outside Time* (3 items)
 - A variety of equipment and materials are available for children to exercise their large muscles and explore and learn from the environment
 - Children are involved in self-directed play during outside time
 - Adults take advantage of opportunities for incidental teaching
- *Circle Time* (3 items)
 - Adults have specific roles in the circle activity
 - Circle activities allow children to get involved in some way and have input
 - Adults take advantage of opportunities for incidental teaching
- *Teacher/Child Interactions* (14 items)
 - Adults extend children's activities and problem-solving by introducing new material
 - Adults extend children's activities and problem-solving by incorporating representation
 - Adults extend children's dramatic play by joining in
 - Adults help children with basic reading and writing skills when children show an interest
 - Adults model new possibilities by playing alongside children
 - Adults make specific comments that extend children's thinking and language and the focus on key experiences
 - Adults extend children's activities and problem solving by making suggestions
 - Adults extend children's activities and problem-solving by asking open-ended and thought provoking questions
 - Adults refer children's questions and comments to other children
 - Adults help children compare number and amount in a functional way
 - Adults expect children to do things for themselves when possible
 - Adults model appropriate communication techniques
 - There is a balance between teacher talk and child talk throughout the day
 - Adults provide children with suggestions for coping with their feelings
- *Classroom Management and Organization* (5 items)
 - Adults turn inappropriate behavior into a problem-solving situation
 - Adults interact with an individual child or small group while maintaining awareness of classroom
 - Adults minimize time spent waiting
 - Adults set reasonable limits, explaining them, and maintaining them
 - Adults use positive guidance techniques
- *Team Evaluation and Planning* (2 items)
 - The classroom staff meets daily to plan and evaluate the day's activities as well as to evaluate the children's progress. The process focuses on the key experiences.
 - Adults use a naturalistic observation method to measure each child's progress in relation to the key experiences. Adults are asked if they use a

record-keeping system for evaluating children's progress. If a record-keeping system is used, adults are asked how they use the information.

II. Administration of Measure

Who administers Measure/Training Required

Test Administration: Information on test administration was not available in materials reviewed.

Training Required: Training on the PCI can take anywhere from three days to two weeks. Training should be ongoing while the observer is using the PCI.

Setting

The PCI is designed for use in programs for children ages three through six which use the High/Scope curriculum. However, the PCI-CD can be used in all pre-school settings regardless of curriculum.

Time Needed and Cost

Time: The authors suggest that observers spend at least one full day in each classroom and observe the classrooms multiple times over the course of a year. Observations should not be made during the first six weeks of the school year, nor prior to or immediately after a holiday.

Cost: Information not available.

III. Functioning of Measure

Reliability Information

Internal Consistency

The Cronbach's alpha of this scale is .89.

Validity Information

Concurrent Validity

The PCI has been found to be significantly correlated with the ECERS-R ($r = .60, p < .01$).

References and Additional Resources

Frede, E. C. & Miller, A. K. (1990). *Preschool Classroom Implementation Rating Instrument – High/Scope Manual*.

Preschool Mental Health Climate Scale (PMHCS)

I. Background Information

Author/Source

Source: Gilliam, W. S. (2008). *Development of the Preschool Mental Health Climate Scale: Final Report*. Yale Child Study Center, New Haven, CT. (unpublished).

Publisher: This measure is currently unpublished.

Purpose of Measure

The purpose of this measure is to evaluate the mental health climate of Head Start and other pre-school classrooms. According to the developer, "None of the existing measures of child care quality were developed to address the full range of classroom characteristics associated with mentally healthy environments for young children – the primary goal of most early childhood mental health consultation" (Gilliam, 2008, p.1). Gilliam (2008) suggests that a reliable and valid measure "would help orient services in this area and could lead to instrumentation more likely to show the positive effects of these interventions," (Gilliam, 2008, p. 2). Further, Gilliam (2008) hopes that this measure will provide "much needed instrumentation for early childhood mental health consultants" (Gilliam, 2008, p. 2).

Two strategies were used to develop an initial pool of 109 pilot items. The first strategy involved generation of items resulting from observational narratives by behavioral consultants in Head Start classrooms, resulting in 69 items. The second strategy involved a review of existing research on early childhood education and child care quality and items in extant measures of child care quality that appeared to be related to mental health and social-emotional development, resulting in 40 items" (Gilliam, 2008, p. 2).

The final measure has 59 items that are scored on a 5-point Likert scale with "1" indicating never or not true, "3" indicating moderately frequent or moderately true and "5" indicating consistently or completely true (Gilliam 2008, p. 6).

Population Measure Developed With

The pilot study took place in 92 early childhood classrooms in Connecticut. There is no demographic information about the children or teachers/staff used in developing the measure. There were five mental health consultants who observed in classrooms to narrow down the item pool. Each consultant had several years of experience and the group had a variety of professional experience.



Age Range/Setting Intended For

The PMHCS is intended for pre-school-age children in Head Start or other pre-school programs.

Ways in which Measure Addresses Diversity

Information not available.

Key Constructs & Scoring of Measure

The measure is divided into Positive Indicators (50 items) and Negative Indicators (9 items).

The positive items are grouped in ten domains:

- A. Transitions
- B. Directions and Rules
- C. Staff Awareness
- D. Staff Affect
- E. Staff Cooperation
- F. Teaching Feelings and Problem-Solving
- G. Individualized and Developmentally Appropriate Pedagogy
- H. Child Interactions.

The negative items are not grouped in domains.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Information not available.

Training Required: Information not available.

Setting

Head Start or other pre-school program classrooms.

Time Needed and Cost

Time: Information not available.

Cost: Information not available.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Inter-rater reliability was established using data from 24 of the pilot classrooms, where two raters independently rated each classroom. Overall, the correlation for Total Positive Indicators was 0.71 and 0.75 for Total Negative Indicators, both falling in the acceptable category as stated by the developers. However, the inter-rater



reliability across the ten domains of positive indicators varied from 0.23 to 0.94, suggesting that some of the domains need "clearer scoring criteria," (Gilliam, 2008, p. 5).

Internal Consistency

Internal consistency reliability was established using data from all 92 classrooms in the pilot study. Internal consistency reliability, recorded as a Cronbach's alpha, for the Total Positive Indicators was 0.98, which the developers consider to be very strong, and 0.75 for the Total Negative Indicators, which the developers consider to be acceptable.

Validity Information

Convergent & Discriminant Validity

Convergent validity was established by correlating the PMHCS with the Arnett Caregiver Interaction Scale (CIS) and some of the domains from the Early Childhood Environmental Rating Scale- Revised (ECERS-R), both of which are well-validated measures in the field. Overall, the correlations suggest a "theoretically consistent relationship" with the other two measures (Gilliam 2008, p. 5). The strongest correlations were between the Directions and Rules, Staff Affect, Staff-Child Interactions, and Individuated and Appropriate Pedagogy domains in the PMHCS with the Positive Interactions Subscale on the CIS; all correlations here were above 0.70. When compared with the ECERS-R, the highest correlations were between the Interactions domain on the ECERS-R and the Staff Awareness and Staff-Child Interactions ($r = 0.73$) and the Total Positive Indicators Score ($r = 0.79$).

References and Additional Resources

Gilliam, Walter S. *Development of the Preschool Mental Health Climate Scale: Final Report*, March 14, 2008 (unpublished).

Preschool Program Quality Assessment, 2nd Edition (PQA)

I. Background Information

Author/Source

Source: High/Scope Educational Research Foundation. (2003). *Preschool Program Quality Assessment, 2nd Edition (PQA) Administration Manual*. High/Scope Press: Ypsilanti, MI.

Publisher: High/Scope Press. A division of the High/Scope Educational Research Foundation. Ypsilanti, Michigan.
www.highscope.org

Purpose of Measure

As described by the authors:

"The Preschool Program Quality Assessment (PQA), Second Edition, is a rating instrument designed to evaluate the quality of early childhood programs and identify staff training needs. Developed by High/Scope Educational Research Foundation, it is appropriate for use in all center-based settings, not just those using the High/Scope educational approach. The Preschool PQA intentionally reflects 'best practices' in early childhood education as a whole. The measure identifies the structural characteristics and dynamic relationships that effectively promote the development of young children, encourage the involvement of families and communities, and create supportive working environments for staff" (High/Scope Educational Research Foundation, 2003, p. 1).

The PQA can be used for a variety of purposes including both pre-service and in-service training initiatives, self-assessment and monitoring. The PQA can also be used to conduct observations and provide feedback to staff. In addition, the Preschool PQA can be used as a research tool when administered by trained outside observers to document program practices, compare quality, examine the relationship between quality of care and children's outcomes, and evaluate the effectiveness of staff development initiatives. Finally, the Preschool PQA can be used to explain research-based practices to a variety of individuals and agencies including administrators, policymakers, and support staff in the pre-school (High/Scope Educational Research Foundation, 2003).

Population Measure Developed With

The current version of the PQA is a revision of earlier versions of the PQA. There are two notable differences: 1) the number of content areas has increased from four to seven, and 2) the scoring system has been revised to adequately measure the full range of quality along each quality construct.

The revised PQA was field tested in two research projects: the 2000 cohort of Phase 2 of the Michigan School Readiness Program (MSRP) evaluation with a sample of 19 classrooms and 2,000 children (Smith, Jurkiewicz, & Xiang, 2002), and the Michigan Full-Day Preschool Comparison Study with two cohorts comprising 121 and 132 classrooms (Jurkiewicz, 2003). A broad range of public and private early childhood settings were represented by these samples, permitting rigorous testing of the psychometric properties of the new PQA.

Age Range/Setting Intended For

The PQA is appropriate for use in all pre-school settings, regardless of whether the center is using the High/Scope educational approach.

Ways in which Measure Addresses Diversity

One item in Section I, the "Learning Environment," rates the extent to which materials in the classroom "reflect human diversity and the positive aspects of children's homes and community cultures." Raters note the extent to which materials reflect the home and community cultures, special needs of children in the program, and a wide range of non-stereotyped role models and cultures. Raters also note the extent to which multicultural materials are integrated into the classroom.

Key Constructs & Scoring of Measure

The PQA is comprised of seven areas of program quality, three of which are based on classroom observation, and four of which are based on interviews with teachers and/or directors. The first four areas are classroom-specific, while the latter three are program-specific. Each area has between 5 and 13 items, with several indicators per item. Raters score each indicator on a 5-point scale. The administration manual provides a detailed description of the scoring procedures. The areas of program quality and items are summarized below.

Classroom Items

- *Learning Environment* (9 items)
 - Safe and healthy environment
 - Defined interest areas
 - Logically located interest areas
 - Outdoor space, equipment, materials
 - Organization and labeling of materials
 - Varied and open-ended materials
 - Plentiful materials
 - Diversity-related materials
 - Displays of child-initiated work
- *Daily Routine* (12 items)
 - Consistent daily routine
 - Parts of the day
 - Appropriate time for each part of day
 - Time for child planning
 - Time for child-initiated activities

- Time for child recall
- Small-group time
- Large-group time
- Choices during transition times
- Cleanup time with reasonable choices
- Snack or meal time
- Outside time
- *Adult-Child Interaction* (13 items)
 - Meeting basic physical needs
 - Handling separation from home
 - Warm and caring atmosphere
 - Support for child communication
 - Support for non-English speakers
 - Adults as partners in play
 - Encouragement of child initiatives
 - Support for child learning at group times
 - Opportunities for child exploration
 - Acknowledgement of child efforts
 - Encouragement for peer interactions
 - Independent problem solving
 - Conflict resolution
- *Curriculum Planning and Assessment* (5 items)
 - Curriculum model
 - Team teaching
 - Comprehensive child records
 - Anecdotal note taking by staff
 - Use of child observation measure

Agency Items

- *Parent Involvement and Family Services* (10 items)
 - Opportunities for involvement
 - Parents on policy-making committees
 - Parent participation in child activities
 - Sharing of curriculum information
 - Staff-parent informal interactions
 - Extending learning at home
 - Formal meetings with parents
 - Diagnostic/special education services
 - Service referrals as needed
 - Transition to kindergarten
- *Staff Qualifications and Staff Development* (7 items)
 - Program director background
 - Instructional staff background
 - Support staff orientation and supervision

- Ongoing professional development
- In-service training content and methods
- Observation and feedback
- Professional organization affiliation
- *Program Management (7 items)*
 - Program licensed
 - Continuity in instructional staff
 - Program assessment
 - Recruitment and enrollment plan
 - Operating policies and procedures
 - Accessibility for those with disabilities
 - Adequacy of program funding

Comments

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The measure may be administered by independent raters including researchers, program evaluators, outside consultants or agency administrators. In addition, site staff including directors, early childhood specialists, curriculum coordinators, teachers, or parents may also complete it as part of a self-assessment. Students may also use their PQA observations as part of their training to become teachers or caregivers.

Training Required: Training to acceptable levels of inter-rater reliability on the PQA takes 2 days. The first day is devoted to reviewing and practicing the PQA, using anecdotes and raw-footage videotapes. The second day is used to conduct actual observations and determine inter-rater reliability.

Setting

The PQA is administered in pre-school classrooms.

Time Needed and Cost

Time: "It is recommended that raters spend at least one full day reviewing a program before completing PQA ratings, allocating a half-day to observing in the classroom (first three sections) and a half-day to conducting interviews (last four sections). If more than one classroom in a center is to be rated, the rater should visit each classroom for a half-day to complete the observations sections and to interview the head teacher" (High/Scope Educational Research Foundation, 2003, p. 5).

Cost: The cost of the PQA is \$25.95.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Pairs of raters were sent to 10 classrooms to observe the learning environment, daily routine, and adult-child interaction. Pearson's correlations were calculated to be 0.57 for learning environment ($p < 0.10$), 0.75 for daily routine ($p < 0.05$), and 0.74 for adult-child interaction ($p < 0.05$).

Internal Consistency

"To assess internal consistency, Cronbach's alpha was calculated on five quality constructs (learning environment, daily routine, adult-child interaction, curriculum planning and assessment) and total PQA scores. There was insufficient data to determine internal consistency on the other two constructs (staff qualifications and development, and program management) since these were only rated once at the agency level rather than for each classroom. . . Internal consistency for the new version was calculated with 185 classrooms in three samples. . . and averaged 0.93, with all but two of the results within the acceptable range of 0.70 to 0.90" (High/Scope Educational Research Foundation, 2003, p. 11).

Validity Information

Concurrent Validity

"The validity of quality constructs within sections I through V of the revised PQA was assessed in relationship to the Teacher Beliefs Scale. . . The PQA was significantly correlated, in the expected positive or negative direction, with appropriate and inappropriate teacher beliefs and practices. With one exception [(the correlation between the learning environment of the PQA and appropriate practices of the Teacher Beliefs Scale, $r = 0.16$)], all correlations were significant and ranged in magnitude from 0.28 to 0.49" (High/Scope Educational Research Foundation, 2003, p. 12).

Predictive Validity

PQA scores are significantly related to children's developmental outcomes, both while children are in pre-school, and kindergarten, and is associated with established measures of child development (e.g. DIAL-R, High/Scope COR) and teacher ratings.

Confirmatory Factor Analysis

"A confirmatory factor analysis was conducted with sections I through V using a sample of approximately 150 classrooms. . . Five factors emerged, accounting for 58% of the variance, and their content aligned with the five corresponding PQA sections: Learning Environment, Daily Routine, Adult-Child Interaction, Curriculum Planning and Assessment, and Parent Involvement and Family Services. Factor loadings ranged from 0.43 to 0.82, with the majority (64%) at 0.60 or higher. However, several daily routine items, notably those related to group times (e.g., small- and large-group time), loaded on the adult-child factor. These items were modified in the

final version of the PQA" (High/Scope Educational Research Foundation, 2003, p. 12).

References and Additional Resources

High/Scope Educational Research Foundation. (2003). *Preschool Program Quality Assessment, 2nd Edition (PQA) Administration Manual*. High/Scope Press: Ypsilanti, MI.

Jurkiewicz, T. (2003). *The Revised Preschool PQA: Report on psychometric properties*. Instrument evaluation report to the Michigan Department of Education. Ypsilanti, MI: High/Scope Educational Research Foundation, Research Division.

Smith, C., Jurkiewicz, T., & Xiang, Z. P. (2002). *Program quality in Michigan School Readiness Program classrooms: Classroom characteristics, teacher beliefs, and measurement issues*. Evaluation report to the Michigan Department of Education. Ypsilanti, MI: High/Scope Educational Research Foundation, Research Division.

Preschool Rating Instrument for Science and Math (PRISM)

I. Background Information

Author/Source

Source: Stevenson-Boyd, J., Brenneman, K., Frede, E., Weber, M. (2008). *Preschool Rating Instrument for Science and Mathematics*. New Brunswick, NJ: National Institute for Early Education Research.

Publisher: This measure is currently unpublished. Please contact Dr. Ellen Frede for further information (efrede@nieer.org).

Purpose of Measure

Developers of the PRISM note a lack of instrument options for assessing instructional supports for early mathematics and science. The PRISM is designed to “assess differences in classroom supports for mathematics and science” (Brenneman, Frede, Stevenson-Boyd, 2009).

The PRISM is based upon the Preschool Classroom Mathematics Instrument (PCMI; Frede, Weber, Hornbeck, Stevenson-Boyd, & Colon, 2005). Both instruments are informed by the National Association for the Education of Young Children and the National Council of Teachers of Mathematics (NAEYC & NCTM, 2002) standards for early mathematics (Brenneman, Frede, Stevenson-Boyd, 2009). The PRISM measures the presence of classroom materials and teaching strategies that support early mathematical and science concept development (Brenneman, Frede, Stevenson-Boyd, 2009). The developers note that the measure includes both the math and science domains because they are conceptually similar (e.g., reasoning that supports classification, seriation, identifying patterns, measurement, and data collection and representation) (Brenneman, Frede, Stevenson-Boyd, 2009).

Population Measure Developed With

The PRISM was developed in publicly supported pre-school classrooms in New Jersey. These include public school, child care and Head Start programs. The instrument is also currently being used in a statewide study of pre-school effects in New Mexico and in a study comparing the effects of English speaking versus Spanish speaking teachers on learning.

Age Range/Setting Intended For

The PRISM is designed for use in pre-school classrooms.

Ways in which Measure Addresses Diversity

There is no explicit attempt to address diversity in the measure.



Key Constructs & Scoring of Measure

The PRISM contains 16 items total. Six items assess the types of materials in the classroom (e.g., materials for counting, comparing, estimating and recognizing number symbols), and 10 items assess staff interactions (e.g., recording science information).

The PRISM contains 11 items that focus on math materials and teacher-child interactions around mathematics concepts, including the following concepts:

- *Supports for counting, comparing, estimating, and recognizing number symbols*
- *Measurement*
- *Classifying and seriating*
- *Geometric thinking and spatial relations*

The PRISM contains 5 science items that focus on materials and teacher-child interactions that support:

- *Explorations of biological and non-biological science*
- *Encouraging reading, writing, and drawing about science*
- *Encourage investigations and discussions of scientific concepts*
- *Support observing, predicting, comparing, and contrasting” (Brenneman, Frede, Stevenson-Boyd, 2009)*

Each item is rated on a Likert scale from 1 to 7, with example indicators (i.e. anchors) given at the score points 1, 3, 5, and 7. A score of 1 indicates the absence of a certain material or practice, and a score of 7 indicates that the material or practice is observed in a close to ideal form.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The PRISM is designed to be used by researchers.

Training Required: At this time researchers must be trained to reliability by the developers. Costs would vary but observers should be knowledgeable in early childhood education and training takes a minimum of 5 days. For further details contact Dr. Ellen Frede at efrede@nieer.org.

Setting

The PRISM is administered in pre-school classrooms.

Time Needed and Cost

Time: The observation period should be an entire half day program and approximately 4 hours in a full day classroom beginning before the children arrive.

Cost: The PRISM is not currently publicly available.



III. Functioning of Measure

Reliability Information

Not yet available.

Validity Information

Content Validity

The math items are based on the NAEYC/NCTM (2002) standards for early mathematics.

The science items were developed based on a review of state early learning standards, science curricula (e.g., French, 2004; Gelman, Brenneman, Macdonald and Roman, in press), and other relevant research (e.g., Chase & Buffton, 2008; Tu, 2006) (Brenneman, Frede, Stevenson-Boyd, 2009).

Comments

The PRISM is based on an earlier measure, the Preschool Classroom Mathematics Instrument (PCMI; Frede, Weber, Hornbeck, Stevenson Boyd & Colon, 2005). The PRISM supersedes the PCMI.

Developers of the PRISM have designed a complementary instrument to be used for systematic professional development that involves self-assessment and mentor-coaching. "The Self-Evaluation of Science and Math Education (SESAME: Frede, Stevenson-Boyd, & Brenneman, 2009) includes criteria for self-assessment that complement the PRISM but that call for teachers and teacher-coaches to provide specific evidence that a particular criterion has been met. The teacher and coach review the results and set specific objectives for improvement" (Frede, Stevenson-Boyd, & Brenneman, 2009, p. 1).

References and Additional Resources

- Brenneman, K., Frede, E., & Stevenson-Boyd, J. (2009) *Preschool Rating Instrument for Science and Mathematics (PRISM) Overview and description*. New Brunswick, NJ: National Institute for Early Education Research.
- Chase, S. & Bluffont, S. (2008). *Model integrity checklist: Science outcomes program classroom review and teacher interview, administrator interview, and supervisor interview*. Pittsburgh, PA: Author.
- Frede, E., Stevenson-Boyd, J., & Brenneman, K. (2009). *Self-Evaluation for Science and Math Education*. New Brunswick, NJ: National Institute for Early Education Research.
- Frede, E., Weber, M., Hornbeck, A., Stevenson-Boyd, J., & Colón, A. (2005). *Prekindergarten Classroom Mathematic Inventory (PCMI)*. New Brunswick, NJ: National Institute for Early Education Research.



- French, L. (2004). Science as the center of a coherent, integrated, early childhood curriculum. *Early Childhood Research Quarterly, 19*, 138-149.
- Gelman, R., Brenneman, K., Macdonald, G., & Roman, M. (in press). *Preschool pathways to science (PrePS): Facilitating scientific ways of thinking, talking, working and knowing*. Brookes Publishing.
- National Association for the Education of Young Children and the National Council of Teachers of Mathematics. (2002). *Early childhood mathematics: Promoting good beginnings*. A joint position statement of the National Association for the Education of Young Children (NAEYC) and the National Council for Teachers of Mathematics (NCTM). Retrieved December 12, 2008 from <http://www.nctm.org/about/content.aspx?id=6352> .
- Stevenson-Boyd, J. S., Brenneman, K., & Frede, E. (2008). *Preschool Rating Instrument for Science and Mathematics (PRISM)*. New Brunswick, NJ: Author.
- Stevenson-Boyd, J., Brenneman, K., Frede, E., Weber, M. (2008). *Preschool Rating Instrument for Science and Mathematics*. New Brunswick, NJ: National Institute for Early Education Research.
- Tu, T. (2006). Preschool science environment: What is available in a preschool classroom? *Early Childhood Education Journal, 33*(4), 245-251.

Quality of Early Childhood Care Settings: Caregiver Rating Scale (QUEST)

I. Background Information

Author/Source

Source: Goodson, B. D., Layzer, J. I., & Layzer, C. J. (2005) *Quality of Early Childhood Care Settings: Caregiver Rating Scale (QUEST)*. Abt Associates Inc.: Cambridge, MA.

Publisher: Abt Associates Inc.
55 Wheeler Street
Cambridge, MA

Purpose of Measure

As described by authors:

"The Caregiver Rating Scale is based on the most up-to-date research on practices that are associated with children's development and learning. The rating scale focuses on caregiver warmth/responsiveness and on caregiver support for the child's development in four important areas—cognitive development, especially language development and early literacy; emotional development; social development; and physical development" (Goodson, Layzer, & Layzer, 2005, p. 5-1).

Population Measure Developed With

The QUEST was developed for use in the National Study of Child Care for Low-Income Families. A major component of this study was a longitudinal study of 650 families using family child care for their children aged one to nine years, and of the family child care providers themselves. Because the study was intended to include a large number of informal providers as well as children across a wide age-range, and followed children when they moved into center-based settings, the developers found no existing measures that were suitable for use across settings.

Age range/Setting Intended For

This measure was intended for use in a variety of settings from informal care to formal center-based care for children 0 to 5 years of age.

Ways in which measure addresses diversity

The QUEST includes 3 items that ask about the caregiver's support for English language learners in the group.

Key Constructs & Scoring of Measure

The current version of the QUEST consists of two measures: the Environment Checklist and the Provider Rating. The Environment Checklist assesses health and safety issues as well as the adequacy and appropriateness of resources in the care environment. The Provider Rating assesses caregiver interactions and behaviors.



The QUEST Environmental Checklist consists of the following subscales:

- *Space and Comfort* (10 items)
- *Equipment and Materials to Support Developmentally-Appropriate Play* (6 items for children < 1; 8 items for children 1 – 3 years; 8 items for children 3 – 5; 7 items for school-aged children)
 - Outdoor Toys and Equipment (1 item for each age group)
- *Equipment and Materials to Support Language and Literacy Development* (12 items)
- *Indoor Safety and Health*
 - Home Furnishings and Materials Equipment (12 items)
 - Exits and Stairs (5 items)
 - Pets (2 items)
- *Daily Routines*
 - Food Preparation, Snack and Meals, Toileting (19 items)
 - Rest Time/Napping (3 items)

Observers rate each item on a scale from 1 (Not true; Little or No Evidence) to 3 (Usually/Always True; Consistent Evidence). Definitions/examples are provided at each scale point for each item.

"The QUEST Caregiver Rating Scale assesses six main aspects of caregiver behavior in the classroom" (Goodson et al., 2005, p. 5-1).

- *The Caregiver with Children*
 - Caring and responding (10 items)
 - Using positive guidance and discipline (9 items)
 - Supervision (4 items)
 - Does no harm (5 items)
- *Supporting Social-Emotional Development* (8 items)
- *Supporting Play* (4 items)
- *Supporting Cognitive Development*
 - Instructional style (5 items)
 - Learning activities and opportunities (11 items)
- *Supporting Language Development and Early Literacy* (11 items)
- *Television and Computers* (2 items)

"The recommended procedure for completing the scale involves three steps: First, the observer collects data on the caregiver's behavior over the entire observation period but only completes the ratings provisionally as additional relevant evidence is observed. Second, at the end of the entire observation period, the observer reviews the provisional codes, revising as needed, and selects a final rating for each code. . Third, in the final step in the coding, the observer completes the nine summary ratings at the end of the rating scale" (Goodson et al., 2005, p. 5-2).



Observers rate each item on a scale from 1 (Not True; Rarely True; Little/No Evidence) to 3 (Usually/Always True; Consistent Evidence). Definitions/examples are provided at each scale point for each item.

II. Administration of Measure:

Who Administers Measure/ Training Required

Test Administration: Trained researchers or investigators administer the measure.

Training Required: Each of the two measures requires a day of training with an additional half-day introduction to the battery and how it should be administered in specific settings.

Setting

For the National Study of Child Care for Low-Income Families the measures were used across a variety of settings from grandmother care to formal center-based care.

Time Needed and Cost

Time needed: Each of the two measures requires a day of training with an additional half-day introduction to the battery and how it should be administered in specific settings.

Cost: The cost of the materials is the cost of reproducing the measures and training manuals. The cost of training is \$2500 for up to 10 trainees, plus expenses.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Paper and pencil tests of observer reliability achieved 85% agreement or better on individual items.

Validity Information

No information is available to date on the validity of the QUEST measure, although two studies have used the QUEST alongside the ECERS and the FDCERS, which will be the basis for validity analyses.

References and Additional Resources

Goodson, B. D., Layzer, J. I., & Layzer, C. J. (2005) *Quality of Early Childhood Care Settings: Caregiver Rating Scale (QUEST)*. Abt Associates Inc.: Cambridge, MA.

Ramey and Ramey Observation of Learning Essentials (ROLE)

I. Background Information

Author/Source

Source: Ramey, S. L. & Ramey, C. T. (2002). *The Ramey Observation of Learning Essentials (ROLE)*, Washington, DC: Georgetown University Center for Health and Education.

Publisher: This measure is currently unpublished.

Purpose of Measure

As described by the authors:

"The ROLE is a quantitative observational tool that describes the type of classroom management, mentoring in basic skills, exploration, and development that the teacher is providing to students in the classroom on a minute-by-minute basis" (Ramey & Ramey, 2002, p. 1). It is based on a comprehensive review of scientific findings that identified seven "essential" environmental transactions that are associated with higher levels of learning and academic achievement in school (Ramey & Ramey, 1999a; Ramey & Ramey, 1999b).

Population Measure Developed With

The ROLE was developed and then used in 40 pre-K and Head Start classrooms in Montgomery County, Maryland's public schools.

Age Range/Setting Intended For

The ROLE is appropriate for use in Pre-Kindergarten classrooms serving 3- and 4-year-old children.

Ways in which Measure Addresses Diversity

The ROLE does not specifically address diversity in the classroom.

Key Constructs & Scoring of Measure

"Observations are conducted throughout the whole classroom day in 5 minute-by-minute intervals, with 2 minutes in between to write notes. During each of the 5 minutes of observation, the observer codes the teacher-child and the paraeducator-child interactions. Each minute's entry includes a code of: the primary learning context; the presence of the teacher, paraeducator, and/or other adult (designated); and the content of the adult-child interaction, if any, in terms of science-based learning essentials. There also are codes for other adult behavior that do not relate to child transactions, such as cleaning/organizing, administration, and non-class-related activities. During the 2-minute interval, the observer records general information about: the number of adults and children present, the content and type of



educational and other activities; and rates the emotional tone in the classroom" (Ramey & Ramey, 2002, p. 1).

- *Instructional value.* This is a rating of the quality of instruction and is measured on a 6 point scale, where 0 indicates "no instruction" and a 5 indicates "excellent instruction."
- *Emotional tone.* "Emotional tone measures the entire classroom climate for the observed 5 minutes, with a focus on the teacher's tone with the students" (Ramey & Ramey, 2002, p. 2). Scores range from 1 to 4, with a 1 indicating "very negative," and a 4 indicating "highly positive."

In addition to the two constructs listed above, minute-by-minute ratings of fifteen teacher-child interactions are also assessed. Observers rate each activity with an "I" if the interaction between the teacher/paraeducator and the child is addressed at only one child and there is no other learning opportunity for the surrounding children, and an "X" if it is a large group or small group interaction. If the activity is not present, the code for the specific interaction is left blank. Brief descriptions of the fifteen activities/opportunities for interaction are presented below.

- *Encourage exploration.* The teacher actively promotes curiosity and exploration of the physical and mental world and the child is encouraged to use senses to independently investigate the world around him/her.
- *Language/literacy/writing.* Working with children on writing, vocabulary, letter sounds, print awareness, reading, and activities which allow children to enhance oral language skills
- *Recreational reading/song/dance.* Reading, singing, or dancing with children that does not introduce any new academic skills or concepts.
- *Math/science/reasoning.* Instruction involving elementary math and science concepts as well as basic reasoning such as sorting, sequence, and patterns.
- *Other instruction.* Includes such activities as instruction involving fine motor skills, arts and crafts and teaching basic hygiene and activities.
- *Celebrate development-specific new skills/academic advancement.* Celebration of a specific, cognitive/intellectual/academic skills or school-related task.
- *Celebrate development—non-academic/general.* General positive affirmations with no academic link.
- *Socialization: guidance and limitation.* Instruction related to the socialization of children to teach which behaviors are acceptable and which are not.
- *Management/monitoring/general supervision.* Phrases used to monitor conduct in a way that does not limit age appropriate behavior, to transition to the next activity, or to check in on students without formal instruction.
- *Unnecessary classroom restrictions.* Unnecessary and inappropriate orders given to children to manage the classroom, eliciting an obedience type of response to a restriction that is inappropriate to the setting and age group.
- *Negative/harsh treatment.* Inappropriate, excessively harsh words or physical treatment of the child.



- *Administrative/cleaning/organizing.* Performing administrative tasks or cleaning and organizing the classroom lasting at least 10 seconds.
- *Conversation with mentor or other school official.* Teacher is communicating with other mentor or other school official.
- *Child assessments.* Teacher is performing formal assessments on children.
- *Non-school related.* Teacher is performing any task that is not related to school.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The test is administered by trained research associates.

Training Required: Extensive training and practice (about 1 to 2 weeks) is required to gain a full grasp of each of the observed activities.

Setting

The ROLE is conducted in the classroom setting of children 3- to 4-years-old.

Time Needed and Cost

Time: Observations are conducted for two to three hours to capture a large range of activities.

Cost: The ROLE is free.

III. Functioning of Measure

Reliability Information

Inter-rater reliability

There are 15 codes/categories on the ROLE (encourage exploration, language/literacy/writing, etc.) and 3 sets of totals that are created for each type of instructional personnel (teacher, instructional assistant, other) creating 45 summary scores per worksheet. Teams of two observers visited 11 classrooms to provide data that would allow inter-rater reliability estimates to be computed. Descriptive statistics were used to explore the degree to which the two raters agreed across the 11 classroom visits. There are a possible 495 cells available for analysis (15 codes/categories x 3 totals x 11 observations).

Raters were in 100% agreement for 320 of the 495 cells (65%). Correlation coefficients between raters across the 11 classroom observations, code/category and instructional personnel range from .91 to 1.00. These correlation coefficients were all statistically significant. There were 175 cells (35%) for which raters' summary scores were different by some frequency between 1 and 13. The majority of the differences were off by a frequency of one ($86/175 = 49\%$). An example of this instance would be one rater counted that the teacher provided language/literacy/writing mentoring 25 times across the entire observation while the



other rater counted that the teacher provided only 24 instances of the same mentoring. Rater discrepancies were deemed large when the frequency difference between raters was greater than or equal to four. Only 40 out of 495 (8%) cells had large discrepancies. These most often reflected codes that occurred with very high frequency throughout the day.

Validity Information

Validity analyses are now underway using longitudinal data and assessments from the ELLCO and curriculum fidelity checklists used in the classrooms. Child outcome data will be related to items and factor scores from research conducted in more than 40 classrooms.

Comments

Future development of the ROLE will involve: (a) an adapted form for use in community-based child care centers and family child care; (b) psychometric research to ascertain the minimum length of time for a ROLE observation session to yield a valid profile of the adult-child transactions throughout the day; (c) the relationship of global classroom rating systems, such as the CLASS (Pianta, La Paro & Hamre, 2008) and the CIS (Arnett, 1989), to this objective, quantitative methodology; and (d) assessment of the usefulness of classroom profiles generated by the ROLE to assist teachers in improving the amount and quality of instructional activities for children.

References and Additional Resources

- Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology*, 10, 541-522.
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System Manual K-3*. Baltimore, MD: Brookes Publishing.
- Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American Psychologist*, 53, 109-120.
- Ramey, C. T., & Ramey, S. L. (1999a). *Right from birth: Building your child's foundation for life*. New York: Goddard Press.
- Ramey, S. L. & Ramey, C. T. (2002). *The Ramey Observation of Learning Essentials (ROLE)*, Washington, DC: Georgetown University Center for Health and Education.
- Ramey, S. L., & Ramey, C. T. (1999b). *Going to school: How to help your child succeed*. New York: Goddard Press.

Ready School Assessment (RSA)

I. Background Information

Author/Source

Source: HighScope Educational Research Foundation. (2006). *Ready School Assessment: Administration Manual*. HighScope Press: Ypsilanti, MI.

HighScope Educational Research Foundation. (2006). *Ready School Assessment: Team Handbook*. HighScope Press: Ypsilanti, MI.

HighScope Educational Research Foundation. (2006). *Ready School Assessment: Questionnaire*. HighScope Press: Ypsilanti, MI.

Publisher: HighScope Press
www.highscope.org
www.readyschoolassessment.net

Purpose of Measure

As described by the authors:

"The focus of the Ready School Assessment (RSA) is on the general policies and practices of a school, with particular emphasis on those that are relevant to the K-2 classrooms, teachers, children, and parents. The RSA is a planning tool designed to provide school improvement teams with a developmental profile of the strength of readiness features in their school" (HighScope Educational Research Foundation, 2006, p. 1).

Population Measure Developed With

Pilot testing of the RSA was conducted in the 2004-2005 school year. For the pilot test, 71 schools in 17 states were recruited and received training. Of these schools, 69 schools from 16 states returned completed data. "Within those 69 schools, 51% identified themselves as urban, 21% as rural, and 25% as suburban (3% defined themselves as "other"). All pilot schools but one were public schools and had an average enrollment of 480 children. Eighty-eight percent of the schools had a prekindergarten program in their building or on the same campus, and 86% of the schools have at least some full-day kindergarten classrooms. In addition, 72.7% of the children in the pilot schools were eligible for free or reduced lunch" (HighScope Educational Research Foundation, 2006, p. 9).

Age Range/Setting Intended For

The measure is intended for elementary schools with an emphasis on pre-K - 2 classrooms, teachers, children, and parents.

Ways in which Measure Addresses Diversity

One of the 8 dimensions measured by the Ready School Assessment is *Respecting Diversity*. This section of the measure contains 20 items. The key constructs assessed in this dimension are: Teaching Diversity, Supporting a Diverse Environment, and Working with Special Needs. The tool defines diversity as class, gender, family background and experiences, and special needs.

Key Constructs & Scoring of Measure

The Ready School Assessment identifies eight major dimensions of what it means to be a ready school. Items are assessed on a scale from "Never" to "Always", Yes and No questions, and numerical frequency questions.

- *Leaders and Leadership* (14 items)
The Principal advocates for and leads the ready school.
 - Principal's commitment
 - Professional Climate
 - Early Childhood Training and Experience
- *Transitions* (18 items)
School staff and parent groups work with families, children, and their pre-school teachers and caregivers before kindergarten and with families and children during kindergarten to smooth the transition from home to school.
 - Transition activities
 - Contact with Pre-K
 - Entry & Promotion
- *Teacher Supports* (11 items)
School organizes classrooms, schedules, teams, and staff activities to maximize the support for all adults to work effectively with children during the school day.
 - Professional Development
 - Contact with Others
- *Engaging Environments* (21 items)
The school's learning environments employ elements that make them warm and inviting, and actively engage children in a variety of learning activities.
 - Safety & Health
 - Materials
 - Classroom Climate
 - Active Learning
- *Effective Curricula* (13 items)
The school diligently employs educational methods/materials shown to be effective in helping children achieve objectives required for grade-level proficiency.
 - Curriculum Training
 - Monitoring Fidelity

- *Family, School, and Community Partnerships* (19 items)
The school takes specific steps to enhance parents' capacities to foster their children's readiness and to support children's learning in and outside of school.
 - Family Involvement in School
 - Parent-School Communication
 - Outreach
- *Respecting Diversity* (20 items)
School helps all children succeed by interacting with children/families in ways that are compatible with individual needs and family backgrounds or life experiences.
 - Teaching Diversity
 - Supporting a Diverse Environment
 - Working with Special Needs
- *Assessing Progress* (13 items)
School staff engage in ongoing improvement based on information that rigorously and systematically assesses classroom experiences, school practices that influence them, and children's progress toward curricular goals.
 - Assessment Mechanisms
 - Using Assessments
 - School Improvement

Comments

A school readiness profile is developed from the ratings of each of the indicators within each dimension. Readiness can be shown for each of the 8 dimensions and the 23 sub-dimensions.

Reports can be automatically generated through the Online Profiler located on the Ready School Assessment website.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: "The RSA is a consensus tool that works best when it brings together a variety of perspectives on the school's readiness. Since the dimensions of the RSA involve many aspects of the school operations, it is best when the assessment is conducted by a team whose members possess a strong understanding of these aspects" (HighScope Educational Research Foundation, 2006, p. 3).

K-2 teachers – the classroom environments and transitional practices

Parents – family involvement practices

Pre-school teachers/Childcare providers – ready school's communication efforts

School Principals – school and district policies, curriculum, and assessment practices

An ideal team consists of four or more members across these groups with knowledge of one or more of the eight dimensions. The ready school team will need to meet

several times to familiarize themselves with the RSA, gather evidence, review/discuss the evidence, and reach consensus on which ready school indicator level best reflects the readiness conditions of their school.

Training Required: "Training on understanding and administering the Ready School Assessment was provided to staff at pilot sites through a series of two-day workshops. Each pilot site was asked to identify a 'ready school team' of at least four or more persons, including (if possible) the school principal, K-2 teachers, pre-school teachers, community leader, parents, and a community-based pre-school or child care staff. Teams participated in the training workshops as groups in order to foster working relationships that would lead to an evidence-based, consensus response to the RSA indicators. The workshop included an introduction to the ready schools concept and provided background for the eight dimensions of school readiness. In addition, the workshops included hands-on practice using the instrument scales as well as practice scoring sessions using school/community case studies taken from the Head Start Transition Study (Love, Logue, Trudeau, & Thayer, 1992)" (HighScope Educational Research Foundation, 2006, p. 9-10).

It is highly recommended that teams complete the initial training piece offered by HighScope Foundation staff. HighScope can provide a full-day of initial assessment training to assist your RSA teams on how best to complete the assessment and action plan for improvement. Training and technical assistance can also be provided in any of the 8 dimensions. This training would be customized to meet the needs of an individual school or district. For more information, or to schedule Ready School Assessment training, please contact HighScope at 734.485.2000, ext 224, via e-mail at infor@highscope.org.

Setting

Observations, anecdotes, parent surveys and teacher surveys are collected about the elementary school environment and operations.

Time Needed and Cost

Time: "The length of time it takes a ready school team to complete the tool will vary depending on the accessibility of the evidence needed to score the tool, the team's ability to work together, and the level of experience the team members have with the self-assessment process" (HighScope Educational Research Foundation, 2006, p. 10).

Cost: \$199.95 + shipping for 5 copies of instrument, 1 Administration Manual, 5 Team Handbooks, 5 Questionnaires, license for online profiler (2 years).

III. Functioning of Measure

Reliability Information

Internal Consistency

Cronbach's alpha was computed for each of the eight RSA dimensions. The alphas ranged from .75 to .93, indicating a high degree of internal consistency. The majority

of alphas (15) for the sub-dimensions were greater than .67. Two sub-dimensions had lower alphas: transition activities (.54) and entry & promotion (.35). Alphas for the dimensions of Leaders and Leadership and Assessing Progress were not presented.

Validity Information

Convergent & Discriminant Validity

"While we might wish for validity based on comparisons of the RSA with other ready school instruments, this possibility is severely constrained by its being essentially the first measurement tool of its kind in the ready school arena" (HighScope Educational Research Foundation, 2006, p. 11).

Content Validity

"The research base for the RSA starts with the attributes of a ready school proposed in the National Education Goals Panel's 1998 report. In developing the RSA instrument, {the author's} carefully reviewed the ready school literature to further flesh out detailed aspects of each of the eight RSA dimensions" (HighScope Educational Research Foundation, 2006, p. 11). In addition, the authors used an advisory panel of elementary school principals, teachers, and early childhood researchers to guide the selection of content and the development of indicators. The instrument was reviewed by focus groups of pre-school program directors, K-2 teachers, and elementary school principals.

"The preliminary review and revision work gave the instrument a strong footing in reality and a good measure of face validity. It also reinforced the content validity derived from its grounding in literature on the ready school topic" (HighScope Educational Research Foundation, 2006, p. 11).

References and Additional Resources

HighScope Educational Research Foundation. (2006). *Ready School Assessment: Administration Manual*. HighScope Press: Ypsilanti, MI.

HighScope Educational Research Foundation. (2006). *Ready School Assessment: Team Handbook*. HighScope Press: Ypsilanti, MI.

HighScope Educational Research Foundation. (2006). *Ready School Assessment: Questionnaire*. HighScope Press: Ypsilanti, MI.

Love, J. M., Logue, M. E., Trudeau, I. V., & Thayer, K. (1992). *Transitions to kindergarten in American schools: Final report of the national transition study*. Washington, DC: US Department of Education.

Shore, R. (1998). *Ready schools*. Washington, DC: National Education Goals Panel.

School – Age Care Environment Rating Scale (SACERS)

I. Background Information

Author/Source

Source: Harms, T., Vineberg Jacobs, E., & Romano White, D. (1996). *School – Age Care Environment Rating Scale*. New York, NY: Teachers College Press.

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

Purpose of Measure

As described by the authors:

The School – Age Care Environment Rating Scale (SACERS) measures environmental quality in school age care settings.

Population Measure Developed With

In order to develop a comprehensive rating scale for school-age child care programs, the authors drew from a number of sources. The SACERS is based on criteria for developmental appropriateness for school-age children.

The SACERS is an adaptation of the Early Childhood Environment Rating Scale (ECERS) (Harms & Clifford, 1980). It is similar in format to the ECERS, the Family Day Care Environment Rating Scale (FDCRS) (Harms & Clifford, 1989) and the Infant/Toddler Environment Rating Scale (ITERS) (Harms, Cryer, & Clifford, 1990), but the content is specific to the school-age care group.

Age Range/Setting Intended For

The SACERS was developed for use with children ages 5- to 12-year-olds.

Ways in which Measure Addresses Diversity

- *Cultural awareness* (item # 27) assesses ethnic, linguistic, gender role, cultural, and racial diversity of toys, books, and pictorial materials. It also assesses encouragement of acceptance and understanding of children with differences as modeled by staff.
- The *Special Needs Supplementary Items* subscale assesses provisions for special needs children. The items in this subscale assess adaptations for children with special needs. The complete list of items in this subscale are presented below.

Key Constructs & Scoring of Measure

Forty-nine items of school-age care environment quality are categorized into seven subscales, each with several items. Items are rated on a 7-point scale from 1 (inadequate) to 7 (excellent). Descriptions are provided at score points 1, 3, 5, and 7.

- *Space and Furnishings* (11 items)
 - Indoor space
 - Space for gross motor activities
 - Space for privacy
 - Room arrangement
 - Furnishings for care routine
 - Furnishings for learning and recreational activities
 - Furnishings for relaxation and comfort
 - Furnishings for gross motor activities
 - Access to host facilities
 - Space to meet personal needs of staff
 - Space to meet professional needs of staff
- *Health and Safety* (8 items)
 - Health policy
 - Health practices
 - Emergency and safety policy
 - Safety practice
 - Attendance
 - Departure
 - Meals/snacks
 - Personal hygiene
- *Activities* (8 items)
 - Arts and crafts
 - Music and movement
 - Blocks and construction
 - Drama/theater
 - Language/reading activities
 - Math/reasoning activities
 - Science/nature activities
 - Cultural awareness
- *Interactions* (9 items)
 - Greeting/departing
 - Staff-child interactions
 - Staff-child communication
 - Staff supervision of children
 - Discipline
 - Peer interactions
 - Interactions between staff and parents
 - Staff interaction

- Relationship between program staff and classroom teachers
- *Program Structure* (4 items)
 - Schedule
 - Free choice
 - Relationship between program staff and program host
 - Use of community resources
- *Staff Development* (3 items)
 - Opportunities for professional growth
 - Staff meetings
 - Supervision and evaluation of staff
- *Special Needs Supplementary Items* (6 items)
 - Provisions for exceptional children
 - Individualization
 - Multiple opportunities for learning and practicing skills
 - Engagement
 - Peer interactions
 - Promoting communication

Comments

There are ~~additional notes~~ "additional notes" for the SACERS that provide more detailed information on specific items that need to be considered while scoring, such as interpretations and explanations of specific wording in the items. These additional notes can be found at the following website: <http://www.fpg.unc.edu/~ecers/>.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The instrument may be used by the caregiving staff for self-assessment, by directors as a program-quality measure for planning program improvement, by agency staff for monitoring, in teacher training programs, and by parents concerned about quality care for their school-age children.

Training Required: Training is required to assure proper use of the instrument for each of its intended uses (i.e., research, program evaluation, and self-evaluation). It is preferable to participate in a training sequence led by an experienced SACERS trainer following the training guide in the SACERS book, pages 38 – 40.

Setting

Observations are made in child care center settings serving children ages 5- to 12-years-old.

Time Needed and Cost

Time: The authors recommend observing for a block of two hours. An additional 20 – 30 minutes is needed to ask the teacher questions on items that were not observed.

Cost: SACERS 1995 \$19.95

The cost of a five-day in-depth training is \$1225/person. A three-day training costs \$825/person. Fees include all materials.

III. Functioning of Measure

Reliability Information

Reliability of the SACERS subscales and total scores was assessed in three ways: internal consistency was calculated using Cronbach's Alphas; inter-rater reliability was measured using the Kappa statistic which corrects for chance agreements; and inter-rater reliability was estimated using intraclass correlations.

Data from 24 after-school programs in two Canadian provinces, Quebec and Ontario, were used to calculate Cronbach's Alphas and Kappas. Two observers independently rated each class on the SACERS during a single visit. One observer rated all 24 classrooms. The second observer was one of five other trained raters. Intraclass correlations require that the same two independent observers rate all groups. These data were available for 13 of the 24 settings. No reliability data was available for on the *Special Needs Supplementary Items* as none of the centers included exceptional children.

Inter-rater Reliability

Weighted Kappas were calculated for 24 centers, rated independently by two observers. Weighted Kappas for each of the subscales and total score are:

<i>Space and Furnishings</i>	.79
<i>Health and Safety</i>	.83
<i>Activities</i>	.86
<i>Interactions</i>	.82
<i>Program Structure</i>	.82
<i>Staff Development</i>	.91
<i>Total Score</i>	.83

Internal Consistency

Cronbach's Alphas for each of the subscales and total scores based on 24 classrooms are:

<i>Space and Furnishings</i>	.76
<i>Health and Safety</i>	.82
<i>Activities</i>	.86
<i>Interactions</i>	.94
<i>Program Structure</i>	.67
<i>Staff Development</i>	.73
<i>Total Score</i>	.95

Intraclass Correlations

Intraclass correlations were calculated on 13 centers that were observed by the same two independent observers. Correlations for each subscale are:

<i>Space and Furnishings</i>	.87
<i>Health and Safety</i>	.95
<i>Activities</i>	.92
<i>Interactions</i>	.93
<i>Program Structure</i>	.99
<i>Staff Development</i>	.99
<i>Total Score</i>	.96

Validity Information

Validity was assessed in two ways: content validity was assessed using expert ratings of each item's importance to their definition of quality; and construct validity was assessed by correlating SACERS total and subscale scores with staff training and staff-to-child ratios.

Construct Validity

SACERS total and subscale scores were correlated with staff training and staff-to-child ratio. Staff training was estimated by assigning a score between 0 and 5 to indicate the highest level of education attained. For example, a score of 5 was assigned if the staff member had completed a college degree in early childhood education or a related field; a score of 4 was given for completion of a college degree in a field unrelated to early childhood education; a score of 3 if the staff member was currently enrolled in an early childhood education or child development program; a score of 2 if the staff member was currently enrolled in a program unrelated to early childhood; a score of 1 for a high school diploma; and a score of 0 if the staff member had not completed high school. Staff-to-child ratios were determined by dividing the total number of children enrolled in the group by the number of staff members assigned to supervise the group. Staff training has moderate positive correlations with Space and Furnishings ($r = .31$), Interactions ($r = .29$), Program Structure ($r = .40$), and Total Score ($r = .29$). Staff-to-child ratios have moderate negative correlations with Health and Safety ($r = -.40$), Activities ($r = -.39$), Staff Development ($r = -.24$), and Total Score ($r = -.30$).

Content Validity

Content validity was assessed by asking nine recognized experts from the United States and Canada to rate the importance of each SACERS item to their intuitive definition of high quality on a 5-point scale (1 = not important to 5 = very important). A mean rating of 4.5 to 5 was found for 91% of the items. The overall mean rating of the items was 4.8. The lowest mean rating assigned to any item was 3.9.

References and Additional Resources

Harms, T., & Clifford, R. M. (1980). *The Early Childhood Environment Rating Scale*. New York, NY.: Teachers College Press.

Harms, T. & Clifford, R.M. (1989). *Family Day Care Rating Scale*. New York, NY: Teachers College Press.

Harms, T., Cryer, D., & Clifford, R. M. (1990). *Infant/Toddler Environment Rating Scale*. New York, NY: Teachers College Press.

Harms, T., Vineberg Jacobs, E., & Romano White, D. (1996). *School – Age Care Environment Rating Scale*. New York, NY: Teachers College Press.

Supports for Early Literacy Assessment (SELA)

I. Background Information

Author/Source

Source: Smith, S., Davidson, S., Weisenfeld, G., & Katsaros, S. (2001). *Supports for Early Literacy Assessment (SELA)*. New York, NY: New York University.

Publisher: This measure is currently unpublished. Contact Dr. Sheila Smith at (646) 284-9600 or Sheila.Smith@nccp.org.

Purpose of Measure

As described by the authors:

The SELA is an instrument still under development that can be used to document the quality of supports for young children's literacy development in center-based pre-school settings. A combination of observations and interview items capture information on both classroom and parent involvement activities.

Some items related to oral language development and developmentally appropriate practice were adapted from the Early Childhood Environment Rating Scale (Harms, Clifford, & Cryer, 1998) and the High/Scope Program Quality Assessment (High/Scope Educational Research Foundation, 2003). Two items related to children learning English as a second language were based on best practice ideas in *One Child, Two Languages* (Tabors, 1997). Other constructs are informed by the NAEYC publication *Learning to Read and Write*.

The SELA is designed for research, training, and professional development efforts to improve the quality of early childhood programs.

Population Measure Developed With

The measure was developed and piloted in pre-kindergarten classrooms serving mostly 4-year-olds in low-income, urban communities. Many classrooms had English language learners and most were ethnically diverse (Hispanic, Asian, African-American children). Programs were publicly funded child care, pre-kindergarten (funded by the state Universal Pre-K program) and/or Head Start. All classrooms were in community settings and many received funds from more than one source (e.g., child care subsidies, Head Start).

Age Range/Setting Intended For

The SELA was developed for use with children ages 3 to 5 in center-based pre-school settings (e.g., child care, pre-kindergarten, Head Start). The instrument is not appropriate for use with younger children.

Ways in which Measure Addresses Diversity

The SELA contains two items for sites with bilingual and non-English speaking children. These items should be used only if 25% of the children in the class are second-language learners.

Key Constructs & Scoring of Measure

A total of 21 rating scales and teacher interview questions represent seven constructs related to early literacy development. Each item within the instrument is rated on a 5-point, Likert-type scale, with a rating of 5 reflecting best practice and a rating of 1 indicating the absence or very low quality of a literacy support. The instrument describes anchor behaviors for ratings of 1 (very low quality), 3 (fair quality), and 5 (ideal quality) for each item.

- *The Literate Environment* (5 items). Assesses the use of environmental print in the classroom, the appearance of the book area in the classroom, the variety of books available to children, the availability of writing materials, and the variety of literacy items and props in the pretend play area.
- *Language Development* (4 items). Assesses whether and how teachers encourage children to use oral language, the richness of teachers' language to children, book reading, and activities that promote children's oral language and knowledge development.
- *Knowledge of Print/Book Concepts* (1 item). Assesses whether the teacher calls attention to the functions and features of print.
- *Phonological Awareness* (1 item). Assesses whether the teacher draws children's attention to the sounds they hear in words.
- *Letters and Words* (2 items). Assesses teachers' promotion of children's interest in writing and the extent to which teachers help children identify letters.
- *Parent Involvement* (2 items). Assesses regular communication between teacher and parent regarding literacy promotion, and special activities to involve parents in their children's literacy development.
- *Developmentally Appropriate Practice* (4 items). Assesses activities and materials, child choice of a variety of developmentally appropriate activities, teacher warmth and acceptance of children, and the promotion of positive interactions among children.
- *Bilingual and non-English speaking children* (2 items). Assesses the extent to which a child's native language is maintained and developed within the classroom setting, and the use of effective strategies to help children understand and acquire English.

Comments

Ratings should reflect what the children are experiencing. That is, if there are multiple teachers in the room, all teacher behavior should be included when determining a rating. Observation notes are the primary source of supporting evidence for ratings, although teacher interview data are also considered. If the

teacher interview indicates a different rating than what was obtained from direct observation, the overall rating on an item can only be elevated (or lowered) one point from the rating based on the direct observation alone.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Raters should be familiar with features of high quality early childhood programs and developmentally appropriate practice.

Training Required: Training on the SELA instrument requires several hours of discussion of the items, achieving inter-rater reliability with a trained rater, and further discussion and resolution of discrepancies in ratings.

Setting

SELA observations occur in center-based settings. Ideally, the observation will include time when staff read to the children.

Time Needed and Cost

Time: Completing the SELA requires a classroom observation of 2.5 to 3 hours and a 30-minute interview with the lead teacher following the observation.

Cost: The instrument is available from the first author, Sheila Smith, Ph.D. (646) 284-9600 or Sheila.Smith@nccp.org.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Lamy et al. (2004) reported that the average inter-rater reliability coefficient for the modified SELA was .98. Raters had to reach at least 80% agreement with an experienced observer on all measures.

In a study conducted in three early childhood settings (Head Start within an elementary school, Head Start in a center-based setting, and pre-kindergarten in a parochial school) in Washington, DC, SELA ratings were obtained for each of four classrooms. However, as with the Lamy et al. (2004) study, this study also used a modified version of the SELA, since some of the classrooms contained children younger than 3. For each classroom observation, ratings were completed by two observers separately, and consensus ratings were arrived at by discussion (Halle, Lavelle, Redd, & Zaslow, 2005). Ratings rarely differed by more than one point on the five-point rating scales prior to conferencing.

In a more recent study of a comparison of the effects of two-way immersion (TWI) and monolingual English immersion (EI), Barnett and colleagues (2007) used the

SELA to measure the quality of the pre-school literacy environment and instruction. They reported an inter-rater reliability coefficient of .80.

Internal Consistency

In a study of a random sample of 310 pre-school classrooms in Abbott County New Jersey, Lamy et al. (2004) used a modified version of the SELA that eliminated 5 items that overlapped with the ECERS-R (that is, the modified SELA had 16 instead of 21 items). They reported that the internal consistency among scale items as measured by Cronbach's alpha was excellent (.92).

Validity Information

Criterion Validity

In a study of a random sample of 310 pre-school classrooms in Abbott County New Jersey, Lamy (2004, as cited in Barnett, Yarosz, Thomas, & Blanco, undated) found that the correlation between SELA and ECERS-R total scores was .75.

Comments

All of the psychometric information reported in this profile is based on modified versions of the SELA, rather than the original instrument.

References and Additional Resources

- Barnett, S. W., Yarosz, D. J., Thomas, J., Jung, K., & Blanco, D. (2007) Two-way and monolingual English immersion in preschool education: An experimental comparison. *Early Childhood Research Quarterly, 22*, 277-293.
- Halle, T., Lavelle, B., Redd, Z., & Zaslow, M. (2005). *Final Report of the Mouth Time Formative Evaluation*. Paper prepared for the CityBridge Foundation. Washington, DC: Child Trends.
- Harms, T., Clifford, R., & Cryer, D. (1998). *Early Childhood Environment Rating Scale – Revised*. New York: Teacher's College Press.
- High/Scope Educational Research Foundation (2003). *Preschool Program Quality Assessment, 2nd Edition (PQA) Administration Manual*. Ypsilanti, Michigan.
- Lamy, C. E., Frede, E., Seplocha, H., Ferrar, H., Wiley, L., & Wolock, E. (2004). *Inch by inch, row by row gonna make this garden grow: Classroom quality and language skills in the Abbott Preschool Program*. Rutgers, NJ: National Institute for Early Education Research.
- Neuman, S. B., Copple, C., Bredekamp, S. (2000). *Learning to Read and Write: Developmentally Appropriate Practices for Young Children*. National Association for the Education of Young Children Publication, Washington, DC.
- Smith, S., Davidson, S., Weisenfeld, G., & Katsaros, S. (2001). *Supports for Early Literacy Assessment (SELA)*. New York, NY: New York University.

Tabors, P. O. (1997). *One child, two languages*. Baltimore, MD: Brookes Publishing.

Supports for Social-Emotional Growth Assessment (SSEGA)

I. Background Information

Author/Source

Source: Smith, S. (2004).

Publisher: This measure is currently unpublished. Contact Dr. Sheila Smith at (646) 284-9600 or Sheila.Smith@nyccp.org.

Purpose of Measure

As described by the authors:

The SSEGA is a new, observation-based classroom assessment instrument that can be used to document the strength of key supports for children's social-emotional growth (e.g., space and materials conducive to small group play), classroom routines and activities (e.g., adequate time for peer interaction, unhurried transitions) and teacher behavior (e.g., guidance to help children develop positive peer relationships and skills in conflict resolution).

Population Measure Developed With

The measure was developed and piloted in pre-kindergarten classrooms serving mostly 4-year-olds in low-income, urban communities. Many classrooms had English language learners and most were ethnically diverse (Hispanic, Asian, African-American children). Programs were publicly funded child care, pre-kindergarten (funded by the state Universal Pre-K program) and/or Head Start. All classrooms were in community settings and many received funds from more than one source (e.g., child care/Head Start).

Age Range/Setting Intended For

The SSEGA was developed for use with children ages 3 to 5 in center-based pre-school settings (e.g., child care, pre-kindergarten, Head Start).

Ways in which Measure Addresses Diversity

The SSEGA does not have specific items that address language or ethnic diversity. However, an item related to "supportive teacher-child relationships" emphasizes the teachers' role in talking positively to children about their individual interests and life circumstances. Two other items assess the classroom's capacity to support the social-emotional growth of children with special needs.

Key Constructs & Scoring of Measure

The instrument's 16 items are clustered as follows:

- *General classroom environment and routines* (3 items)
- *Supportive teacher-child relationships* (2 items)
- *Supports for emotional self-regulation* (2 items)

- *Supports for children's positive social behavior* (3 items)
- *Supports for children's social understanding* (2 items)
- *Parent involvement* (2 items)
- *Program identify and support children with special needs* (2 items)

Each item is rated on a 5-point Likert-type scale, with a rating of 5 reflecting best practice and a rating of 1 indicating the absence or very low quality of a support for social-emotional growth. The instrument describes anchor behaviors for ratings of 1 (very low quality), 3 (fair quality), and 5 (ideal quality) for each item.

Comments

Ratings should reflect what the children are experiencing. That is, if there are multiple teachers in the room, all teacher behavior should be considered when determining a rating. Observation notes are the primary source of supporting evidence for ratings, although teacher interview data are also considered. If the teacher interview indicates a different rating for an item than what was obtained from direct observation, the item's final rating can only be elevated (or lowered) one point from the rating based on the direct observation alone.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Raters should be familiar with features of high quality early childhood programs and developmentally appropriate practice.

Training Required: Training on the SSEGA instrument requires several hours of discussion of the items, achieving inter-rater reliability with a trained rater in actual classroom assessments, and further discussion and resolution of discrepancies in ratings. Classroom vignettes have been developed for the initial phase of training to allow practice ratings with the SSEGA.

Setting

The SSEGA is administered in center-based pre-school settings.

Time Needed and Cost

Time: Completing the SSEGA requires a classroom observation of 2.5 to 3 hours and a 30 minute interview with the lead teacher following the observation.

Cost: The instrument is available from the author, Sheila Smith, Ph.D. (646) 284-9600 or Sheila.Smith@nccp.org.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

In an unpublished pilot study (Smith, 2004) raters were consistently able to reach 90% or above agreement within a point after a brief period of training involving practice ratings of the SSEGA based on classroom vignettes. Raters were knowledgeable about high quality early childhood education settings prior to training.

Validity Information

Criterion Validity

In a pilot study of 36 classrooms in New York City, relationships were found between the total SSEGA score, measures of professional development and teachers' views of best practices. For both teachers and assistant teachers, the SSEGA total score was significantly correlated with reports of the number of workshops related to pre-schoolers' social-emotional growth attended in the past year (.59 and .41 respectively), and for assistant teachers, with on-site coaching (.38). For both teachers and assistant teachers, the SSEGA total score was significantly correlated with average ratings of teachers' responses to two social-emotional problem scenarios (.38 and .54 respectively). In these scenarios, teachers described how they would approach different situations in which children need help following classroom routines, managing and understanding emotions and understanding others' intentions. In the coding scheme, higher ratings reflected best practices described in the SSEGA.

Content Validity

The SSEGA was developed through a process of reviewing current research-to-practice guidelines for supporting pre-schoolers' social-emotional growth and refining items based on feedback from leading scholars in this area. (Major sources used in the development of the SSEGA are cited in the instrument.)

References and Additional Resources

Smith, S. (2004) unpublished study. *Pilot of a new classroom assessment instrument: Supports for Social-Emotional Growth*. New York, NY: Child and Family Policy Center, Steinhardt School of Education, Culture, and Human Development, New York University.

Teacher Behavior Rating Scale (TBRS)

I. Background Information

Author/Source

Source: Landry, S. H., Crawford, A., Gunnewig, S., Swank, P. R. (2001). *Teacher Behavior Rating Scale*. Center for Improving the Readiness of Children for Learning and Education, University of Texas Health Science Center at Houston, unpublished research instrument.

Assel, M. A., Landry, S. H., & Swank, P. R. (2008). Are early childhood classrooms preparing children to be school ready?: The CIRCLE Teacher Behavior Rating Scale. In L. Justice & C. Vukelich (Eds.), *Achieving Excellence in Preschool Literacy Instruction*, (pp. 120-135). New York, NY: The Guilford Press.

Publisher: Children's Learning Institute (CLI)
7000 Fannin
Houston, TX 77030
(713) 500-3710
Website: <http://www.childrenslearninginstitute.org/>

Purpose of Measure

As described by the authors:

The Teacher Behavior Rating Scale (TBRS) is an observational tool designed to assess the quantity and quality of general teaching behaviors, language use, and literacy instruction in early childhood classrooms. The TBRS was developed to be "sensitive to classroom environments and instructional practices that promote the skills important for school readiness," as well as "to ensure that the instructional areas measured were predictive of change in children's literacy and language skills, thus providing documentation that improvement in teaching practices would promote improvements in children's academic readiness" (Assel, Landry, & Swank, 2008 p. 123).

Ratings are based on teacher behaviors, child engagement during learning activities, the presence of rich print and learning materials, and evidence that learning activities are planned and responsive to children's needs. The instrument is appropriate for use as both a process and outcome measure, and may be administered by program or research staff.

Population Measure Developed With

The measure was originally developed to provide information to mentors working with Head Start teachers participating in an extensive professional development program, and to evaluate the impact of the professional development on children's

literacy, math, and social skills. Several items were added, and scoring procedures were adjusted over the course of subsequent studies conducted in low-income early childhood classrooms across Texas, Ohio, Florida, and Maryland. The tool has been utilized in urban and rural areas, and in classrooms serving children from families with varied income levels. The measure has been used in the following studies of child care, Head Start, and public pre-k classrooms:

- 106 teachers randomly selected from a cohort of 152 Head Start, Title I, and universal Pre-k classrooms (PCER- Preschool Curriculum Evaluation Research; funded by the Institute of Educational Sciences)
- 75 teachers randomly selected from a cohort of 262 Head Start, child care, and public Pre-k classrooms in four states (IERI study; funded by the Institute of Educational Sciences)
- 161 teachers randomly selected from a cohort of 215 Head Start, child care, and public Pre-k classrooms in 11 Texas communities (Texas Early Education Model; state legislative demonstration early childhood project)
- 92 teachers from Head Start, Title I, and universal Pre-K classrooms in Texas and Florida (Program Project; funded by the National Institute of Child Health and Human Development)

Age Range/Setting Intended For

The pre-K TBRS can be used to observe teachers and caregivers of children 3 to 5 years of age, and can be used in a variety of early care and education settings.

Ways in which Measure Addresses Diversity

TBRS scales for oral language use, book reading, and literacy instruction were adapted for use in bilingual classrooms, allowing observers to capture differences in the frequency and quality of teaching behaviors in multiple languages. Field testing of the Bilingual-TBRS is being conducted in approximately 135 dual language classrooms with teachers speaking Spanish and English, serving English Language Learners.

Key Constructs & Scoring of Measure

The TBRS is comprised of 11 subscales, and 68 items:

Content of Each Construct:

- *Classroom Community (5 items)*
 - Teacher orients child for classroom expectations
 - Children participate in rules and routines
 - Children can move safely around the room
 - Materials are accessible to children
 - Displays children's work around the room
- *Sensitivity Behaviors (12 items)*
 - Sensitive response to children's cognitive signals
 - Encourages children to regulate their behavior
 - Uses non-specific praise and encouragement

- Uses specific praise and encouragement
 - Deepens children's understanding
 - Fails to respond to comments and questions
 - Sensitive response to children's affective signals
 - Response Style varies across children
 - Presence of negative language
 - Presence of positive non-verbal behaviors
 - Presence of negative non-verbal behaviors
 - Uses playful techniques to make cognitive activities engaging
- *Book Reading Behaviors (9 items)*
- Introduces the book
 - Encourages discussion of book features
 - Discusses vocabulary words
 - Vocabulary words are combined with pictures or objects
 - Reads with expression
 - Pace allows children to be involved in read aloud
 - Asks questions to encourage discussion
 - Extends book through activities and discussion
 - Number of children included in read aloud
- *Oral Language Use (7 items)*
- Teacher speaks clearly
 - Models speaking in complete sentences
 - Uses scaffolding language
 - Uses thinking questions
 - Makes links with previously learned words and concepts
 - Encourages language throughout the observation period
 - Engages children in conversations
- *Print and Letter Knowledge (7 items)*
- Promotes letter word knowledge
 - Compares and discusses differences in letter and words
 - Discusses concepts of print
 - Breadth of print and letter activities
 - Literacy connection in centers
 - Print in the environment and centers
 - Letter wall
- *Phonological Awareness (8 items)*
- Integrates PA activities
 - Listening
 - Sentence segmenting
 - Syllable blending and segmenting
 - Onset rime Blending and segmenting

- Rhyming
- Phoneme blending, segmenting, and manipulation
- Alliteration
- *Written Expression (3 items)*
 - Teacher models writing
 - Opportunities to engage in writing
 - Writing materials in centers
- *Math Concepts (5 items)*
 - Hands on math activities
 - Math in daily routines
 - Breadth of math activities
 - Presence of specific math manipulatives
 - Impromptu math teaching moments
- *Centers (6 items)*
 - Centers linked to theme
 - Theme linked materials and activities
 - Prepares children for centers
 - Centers have clear boundaries
 - Centers encourage interaction
 - Models use and care of center activities
- *Lesson Plans (3 items)*
 - Lesson plan linked to theme
 - Implements lesson plan
 - Lesson plan objective are evident
- *Assessments and Portfolios (6 items)*
 - Uses literacy checklists and assessments
 - Uses math checklists and assessments
 - Uses assessments for planning
 - How to use assessments for planning
 - Portfolios of children's progress
 - Uses anecdotal notes
- *Team Teaching (5 items)*
 - Assistant provides classroom instruction
 - Assistant scaffolds language in small group
 - Assistant scaffolds language throughout observation
 - Assistant participates in classroom regulation
 - Assistant improves teaching environment

Most TBR items consist of a 3-point quantity rating (defined as rare, sometimes, often), and a 4-point quality rating (defined as low, medium low, medium high, high).

Frequencies, along with quality ratings, are recorded for several items capturing specific literacy and math instructional moments. Completion of items from Lesson Plans, Assessment and Portfolios subscales require document review and a brief teacher interview.

Comments

A modified version of the TBRS is used when information about the instructional practices of multiple classroom teachers is required. A version of the TBRS appropriate for K – 1st grade is also available.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The TBRS may be administered by trained observers in a variety of applications including:

- Researchers interested in assessing the quantity and quality of early childhood classrooms and instructional practices
- Classroom coaches or mentors gathering information to guide professional development efforts
- Evaluators establishing routine performance measurement, or program results

Training Required: A minimum of 2 days of training is required for use of the TBRS. Day 1 provides trainees with an overview of TBRS subscales to be used, discussion and guidance for using rating scales reliably, and the viewing of several exemplars. Trainees spend day 2 coding and discussing videotaped classroom observations displaying varied instructional quality. At the conclusion of day 2, a set of videotaped practice observations and scoring key is provided to assist in the establishment of good coder agreement. Average subscale agreement within 1 point of master coder scoring is suggested for responsible use of the TBRS. Typical training time for observers without experience in early childhood education is 2 weeks. It is recommended that observers demonstrate agreement within 1 rank of master coders on "field observations" before collecting data independently.

Setting

The measure was developed for use in early childhood classrooms. It has been adapted for use in dual language classrooms, and K-1st grade classrooms.

Time Needed and Cost

Time: Completion of the TBRS requires 2-3 hours of observation time while the teachers of interest are with their children. Observations should be conducted during the time of day when cognitive readiness activities are most likely to occur.

Cost:

Manual: \$35

Scoring Forms: Included with manual

Training DVDs: \$80 (includes train the trainer presentation disc and practice classrooms)
2 days of training at UTHSC-Houston: \$750 x 2 days; 26% indirect= \$1890
2 days of local training: \$750 x 2 days; 26% indirect= \$1890 (plus travel expenses for 1 trainer)

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Generalizability coefficients for inter-rater reliabilities for TBR subscales range from .80 to .98.

Internal Consistency

Internal consistency for the total scale is .96.

Stability across Time

The correlation between total scores for a group of control teachers observed two months apart was 0.75.

Validity Information

Construct Validity

Evidence of convergent validity is seen in multiple instances in which teachers with higher scores on the TBR also have students who score higher on measures of early literacy.

Child Outcomes	General Teaching Behaviors	Oral Language Use	Phonological Awareness	Print and Letter Knowledge
PLS-IV Auditory Comprehension	.54***(.0001)	.61***(.001)	.34***(.0004)	.46***(.0001)
Expressive Vocabulary Test	.57***(.0001)	.63***(.0001)	.37***(.0002)	.53***(.0001)
WJ-III Letter Word ID	.36***(.0002)	.51***(.001)	.25* (.011)	.37***(.0001)
WJ-III Sound Awareness (Rhyming)	.35***(.0003)	.62***(.001)	.39*** (.0001)	.55***(.0001)
DSC Auditory	.40 ***(.0001)	.47***(.001)	.31**(.0017)	.44***(.0001)

Note. * significant at the .05 level.; ** significant at the .01 level; *** significant at the .001 level.

Significant correlations, ranging from .40 to .66, between Bilingual-TBRS subscales and the BESA, EOWPVT, ROWPVT, and tests of phonological awareness and letter knowledge have also been found.

Further evidence of convergent validity was found when correlating TBRS subscales with children’s gains in language and literacy skills. Teachers’ oral language scores correlated .61 and .62 with children’s growth in language and vocabulary, .51 with alphabet knowledge, .47 with phonological awareness.

Additional evidence of construct validity is seen when comparing TBRS scores between target and control teachers. Significant group differences were seen, with target teachers scoring an average of 1.5 points higher on a 1-5 scale. Across multiple waves of observation, target teachers also had a significantly faster rate of growth than control teachers. Similarly, the Bilingual-TBRS shows significant group differences between target and control teachers, in oral language use, general teaching behaviors, and total scale scores.

Criterion Validity

There are few criterion measures available that measure classroom behavior in the detail that the TBRS does. In addition, the purpose of the scale is to assess classroom behavior at the moment and not in the future. Thus, criterion-related validity evidence is not available at the present time.

Concurrent Validity

Significant correlations between TBRS items and teacher self-reports of knowledge and instructional approaches, for a sample of 100 randomly selected teachers, ranged from .25 to .40.

Content Validity

The TBRS was developed to capture what research has shown are the critical components of instruction related to early literacy, language, early math, and responsive teaching.

References and Additional Resources

- Assel, M. A., Landry, S. H., & Swank, P. R. (2008). Are early childhood classrooms preparing children to be school ready?: The CIRCLE Teacher Behavior Rating Scale. In L. Justice & C. Vukelich (Eds.), *Achieving Excellence in Preschool Literacy Instruction*, (pp. 120-135). New York, NY: The Guilford Press.
- Landry, S. H., Crawford, A., Gunnewig, S., Swank, P. R. (2001). *Teacher Behavior Rating Scale*. Center for Improving the Readiness of Children for Learning and Education, University of Texas Health Science Center at Houston, unpublished research instrument.
- Landry, S. H., Swank, P. R., Smith, K. E., Assel, M. A., & Gunnewig, S. (2006). Enhancing early literacy skills for pre-school children: Bringing a professional development model to scale. *Journal of Learning Disabilities*, 39, 306-324.
- Landry, S. H., Anthony, J. L., Swank, P. R., Monseque-Bailey, P. (2009). Effectiveness of Comprehensive Professional Development for Teachers of At-Risk Preschoolers. *Journal of Educational Psychology*, 101, 448-465.
- Landry, S. H., Swank, P. R., Assel, M. A., & Anthony, J. L. (in press). An experimental study evaluating professional development activities within a state funded pre-kindergarten program: Bringing together subsidized childcare, public school, and Head Start.

Teacher Instructional Engagement Scale (TIES)

I. Background Information

Author/Source

Source: Dickinson, D. (2008). *Teacher Instructional Engagement Scale*. Nashville, TN: Vanderbilt University.

Publisher: This measure is currently unpublished. For use, contact David Dickinson at David.Dickinson@Vanderbilt.Edu.

Purpose of Measure

As described by the authors:

The tool seeks to provide a quick description of ways that both the lead and assistant teacher engage children during occasions when research indicates that children may benefit from rich teacher-child conversations. The three included sub-measures (Centers Time Engagement, Teacher Meal Time Engagement, and Quality of Book Reading) value intentional efforts to use and define vocabulary and to engage children in sustained conversations. The tool does not highlight print-based instruction.

Population Measure Developed With

This tool was developed for use in pre-school classrooms that serve low-income children. The scale was based on findings from the Home-School Study of Language and Literacy Development (Dickinson & Tabors, 2001).

Age Range/Setting Intended For

The TIES is designed to assess teachers of pre-school-aged children.

Key Constructs & Scoring of Measure

The tool seeks to describe several features of teacher-child interaction:

- vocabulary instruction
- sustained conversations
- instructional engagement with children related to language or support for using print

There are three sub-measures included in this measure. The items in the first two sub-measures (Centers Time Engagement and Teacher Meal Time Engagement) are rated for both the Lead and Assistant Teacher. The items in the third sub-scale, Quality of Book Reading, are only given one score per classroom. Items are rated as either having been observed or not.

- *Centers Time Engagement* (10 items)
 - Example Item: Spends 4-5 minutes talking in two areas with a focus on activity being done in that area (blocks, art, etc.).

- *Teacher Meal Time Engagement* (10 items)
 - Example Item: Sits with children for most of the observed meal time (3/4 or more).
- *Quality of Book Reading* (11 items)
 - Example Item: Teacher reads in a manner designed to hold attention: varies volume, pace, may use facial expression or gesture.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Trained observers can administer the measure via live classroom observation or video-taped sessions.

Training Required: The developer provided training to the trained observers for the piloting of this measure. It was relatively easy to achieve inter-rater reliability in live coding of pre-school classrooms. No formal training materials have been developed.

Setting

Three settings in pre-school classrooms:

- Centers time (free choice)
- Meal (breakfast or lunch)
- Book reading

Time Needed and Cost

Time: Observations for occur for at least 20 minutes for each of the three settings (minimum total of 60 minutes)

Cost: The measure has not been published and can be obtained free of cost from the developer. To obtain a copy, contact Dr. David Dickinson at David.Dickinson@Vanderbilt.Edu.

III. Functioning of Measure

Reliability Information

Inter-rater Reliability

Setting	% of sessions coded by 2 independent observers	% exact agreement between observers
Book Reading	13.5	93.4
Meals (Lead)	15.4	91.3
Meals (Asst)	15.4	87.5
Centers (Lead)	15.4	88.8
Centers (Asst)	15.4	86.3

Internal Consistency

Setting	Number of Items	Number of Teachers Used for Analysis	Cronbach's Alpha
Book Reading	11	52	.638
Centers	10	103	.656
Meals	10	104	.712

Stability across Time

Information on stability across time is currently unavailable. All three sub-measures were administered only one time. Multiple administrations across the length of the study are recommended.

Validity Information

Concurrent Validity

Pearson Correlations between Teacher Engagement measures and ELLCO subscales

Book Reading Engagement correlated with:

ELLCO – General Classroom Environment subscale ($r = .38, p < .01$)

ELLCO – Language, Literacy, and Curriculum subscale ($r = .32, p < .05$)

Centers (Lead) correlated with:

ELLCO – Literacy Activities Rating Scale ($r = .41, p < .01$)

ELLCO – General Classroom Environment subscale ($r = .34, p < .05$)

ELLCO – Language, Literacy, and Curriculum subscale ($r = .53, p < .01$)

Centers (Asst) correlated with:

ELLCO – Literacy Activities Rating Scale ($r = .38, p < .01$)

Meals (Lead) and Meals (Asst) were not significantly correlated with any ELLCO subscale.

References and Additional Resources

Dickinson, D. (2008). *Teacher Instructional Engagement Scale*. Nashville, TN: Vanderbilt University.

Dickinson, D. K. & Tabors, P. O. (2001). *Beginning literacy with language*. Baltimore, MD: Brookes Publishing.

Teacher Knowledge Assessment (TKA)

I. Background Information

Author/Source

Source: Neuman, S. B. & Cunningham, L. (2009). The impact of professional development coursework and coaching on early language literacy practices. *American Educational Research Journal*, 46(2), 532-566.

Publisher: This measure is currently unpublished. Contact Dr. Susan B. Neuman: sbneuman@umich.edu

Purpose of Measure

As described by the authors:

The Teacher Knowledge Assessment was constructed to examine teachers' knowledge of early language and literacy development and practice. It is a criterion-referenced assessment comprised of multiple-choice and true/false items.

Population Measure Developed With

The measure was developed in collaboration with community college instructors to determine students' growth in knowledge of language and literacy as a result of professional development and/or a formal course in language and literacy development.

Age Range/Setting Intended For

This instrument is intended to assess the language and literacy knowledge of early childhood practitioners or those new to the field. It has been used to assess the knowledge of practicing teachers and caregivers before and after they receive professional development (Koh & Neuman, 2009; Neuman & Cunningham, 2009; Neuman & Wright, in press). It also has the potential to be used during early childhood teacher preparation to determine whether pre-service teachers have acquired knowledge in this area.

Ways in which Measure Addresses Diversity

The instrument examines eight core competencies in early language and literacy development. One of these competencies includes diversity, and sensitivity to cultural differences and perspectives.

Key Constructs & Scoring of Measure

The instrument consists of 45 multiple-choice questions, designed to tap high-quality early language and literacy knowledge and practice, and 25 true-false questions for a total of 70 items.



Forty-eight items were constructed in eight subscales in alignment with the core competencies in Language and Literacy: Oral Language Comprehension; Phonological Awareness; Letter Knowledge/ Alphabetic Principle; Print Conventions; Strategies for Second Language Learners; Literacy Assessment; Parental Involvement; and Literacy Links across the Curriculum. These core competencies reflected accreditation standards of the National Association for the Education of Young Children (NAEYC), the National Association for Family Child Care (NAFCC), the International Reading Association (IRA), and State of Michigan child care licensing requirements.

In addition to the eight subscales assessing language and literacy knowledge, an additional subscale comprised of 22 of the items was included to assess practitioners' foundational knowledge in early childhood development and education. This subscale assessed practitioners' knowledge in the following content areas: Child Development and Learning; Health, Safety and Nutrition; Family and Community Collaboration; Program Management; Teaching and Learning; Observation, Documentation, and Assessment; Interaction and Guidance; and Professionalism. Results indicated that this subscale was significantly correlated with all others ($r = .32$), confirming the theoretical assumption that a strong knowledge base of foundational knowledge in early childhood education is positively associated with a knowledge base in early language and literacy. This subscale also allows for assessment of early childhood foundational knowledge.

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: The instrument is completed by the educator his/herself. It can be administered through SurveyMonkey.com online or through traditional paper and pencil techniques.

Training Required: No training is required.

Setting

The TKA can be used in the field or to determine whether pre-service teachers have acquired knowledge in this area.

Time Needed/Cost

Time: The average completion time is 45 minutes. Two forms of the Assessment were created for pre- and post-test purposes. The test is scored by calculating the total number of correct answers or by calculating a percentage of correct answers.

Cost: Free with request of authors.



III. Functioning of Measure

Reliability Information

Internal Consistency

Results from the administration of the Language and Literacy Assessment with 304 community college students indicate excellent overall reliability ($\alpha = .96$).

Results from a confirmatory factor analysis of the nine subscales (eight in language and literacy, one in early childhood foundations) resulted in only one factor with an Eigen value of 3.198 ($\alpha = .74$), accounting for 36% of the variance. These results indicated that the nine subscales in this assessment worked well together to define a corpus of early language and literacy knowledge that can be accurately assessed by this instrument.

Validity Information

Content Validity

"This assessment was reviewed by several experts in the field of early literacy to ensure that the content was accurate and research based. Each community-college instructor reviewed the assessment for content validity and alignment with course syllabus" (Neuman & Cunningham, 2009, p. 544).

References and Additional Resources

Koh, S. & Neuman, S. B. (2009). The impact of professional development on family child care: a practice-based approach. *Early Education and Development, 20* (3), 537-562.

Neuman, S. B. & Cunningham, L (2009). The impact of professional development coursework and coaching on early language and literacy instructional practices. *American Educational Research Journal, 46*, 532-566.

Neuman, S. B. & Wright, T. S. (in press). Promoting Language and Literacy Development for Early Childhood Educators: A Mixed-Methods Study of Coursework and Coaching. *Elementary School Journal*.

Teaching Pyramid Observation Tool for Preschool Classrooms (TPOT)

I. Background Information

Author/Source

Source: Hemmeter, M. L., Fox, L., & Snyder, P. (2008). *Teaching Pyramid Tool for Preschool Classrooms (TPOT)*. Nashville, TN: Vanderbilt University.

Publisher: This measure is currently unpublished.

Purpose of Measure

"The Teaching Pyramid Tool for Preschool Classrooms (TPOT) provides a tool for assessing the fidelity of implementation of the *Teaching Pyramid* model. Items on the checklist serve as indicators that teaching practices associated with each component of the intervention are in place" (Hemmeter, Fox, & Snyder, 2008, p. 1).

The Teaching Pyramid Model is "a model for supporting social competence and preventing challenging behavior in young children," (Fox, Dunlap, Hemmeter, Joseph, & Strain, 2003, p. 49). There are four levels of this pyramid that guide teacher and child interactions to help social and emotional development. The first level, at the bottom of the pyramid, is to build "positive relationships with children, families, and colleagues." The second level involves implementing "classroom preventative practices", which is followed by "using social and emotional teaching strategies." Finally, at the top of the pyramid, is "planning intensive, individualized interventions" for children when necessary, (Fox et al. 2003, p. 43 - 44).

Population Measure Developed With

Information not available.

Age Range/Setting Intended For

The TPOT is intended for use in pre-school classrooms.

Ways in which Measure Addresses Diversity

Information not available.

Key Constructs & Scoring of Measure

The TPOT contains 22 items. Across these items, there are three types of formats:

1. Items that require a yes/no response based on the observation (items 1 - 7)
2. Items that require a rating based on the observation and teacher interviews (items 8 - 18)
3. Items that are scored based on responses given by the teacher who is observed (items 19 - 22)



The 22 items also are grouped into four main constructs, each of which includes several practices:

- *Responsive Interactions* (13 items)
 - Teachers engage in supportive conversations with children
 - Providing directions
 - Using effective strategies to respond to problem behavior
 - Describe how you communicate with your families and promote family involvement in the classroom
 - Strategies used to build collaborative teaming relationships with other adults
 - Teacher talk to children is primarily giving directions, telling children what to do, reprimanding children
 - Children are reprimanded for engaging in problem behavior (use of "no", "stop", "don't")
 - Children are threatened with an impending negative consequence that will occur if problem behavior persists
 - Teacher reprimands children for expressing their emotions
 - Teacher's guidance or focus around relationships is on adult-child interactions
 - Teacher comments about families are focused on the challenges presented by families and their lack of interest in being involved
 - Teacher only communicates with families when children have challenging behavior
 - Teacher complains about other team members and notes difficulty in their relationships

- *Classroom Preventative Practices* (14 items)
 - Learning centers have clear boundaries
 - The classroom is arranged such that all children in the classroom can move easily around the room
 - The classroom is arranged such that there are no large, wide open spaces where children could run
 - There is an adequate number and variety of centers of interest to children and to support the number of children (at least 4 centers, 1 center per every 4 children)
 - Materials/centers are prepared before children arrive at the center activity
 - Classroom rules or program-wide expectations are posted, illustrated with a picture or photo of each rule or expectation, limited in number (3-5), and stated positively (all have to be true to score a "yes")
 - Schedules and routines
 - Transitions between activities are appropriate
 - Promoting children's engagement



- Teaching children behavior expectations (i.e. posted classroom rules or program wide expectations)
 - The majority of the day is spent in teacher directed activities
 - Many transitions are chaotic
 - During group activities, many children are NOT engaged
 - Teachers are not prepared for activities before the children arrive at the activity
- *Social and Emotional Teaching Strategies* (7 items)
 - Teaching social skills and emotional competencies
 - Teaching children to express emotions
 - Teaching problem solving
 - Supporting friendship skills
 - Emotions are not generally discussed in the classroom
 - Teachers gives group directions to all children in the same way
 - Teacher tells children mostly what not to do rather than what to do
 - *Individualized Interventions* (3 items)
 - Supporting children with persistent problem behavior
 - Involving families in supporting their child's social emotional development and addressing problem behavior
 - Teacher asks for the removal of children with persistent challenging behavior from the classroom or program

II. Administration of Measure

Who Administers Measure/Training Required

Test Administration: Information not available.

Training Required: Information not available.

Setting

The TPOT is intended for use in pre-school classrooms.

Time Needed and Cost

Time: The TPOT is completed during an observation of the classroom and after interviewing the teacher. Observations last at least 2 hours and should include both teacher- and child-directed activities (Fox et al., 2008).

Cost: Information not available.

III. Functioning of Measure

This measure is still undergoing development. Currently, there is a study being conducted to measure the psychometric integrity of the TPOT. Thus, there is no information about the reliability or validity to report at this time.



Comments

This measure should not be used without written permission from the authors, and should not be used for research without a research agreement with the authors (Lisa Fox, fox@fmhi.usf.edu, Mary Louise Hemmeter, ML.Hemmeter@Vanderbilt.edu, and Pat Snyder, particiasnyder@coe.ufl.edu).

References and Additional Resources

Fox, L., Hemmeter, M. L., Snyder, P., Artman, K., Griffin, A., Higgins, A., Kinder, K., Morris, J., Robinson, E., & Shepcaro, J. (2008). *Teaching Pyramid Observation Tool for Preschool Classrooms (TPOT) Manual Research Edition*. Nashville, TN: Vanderbilt University.

Fox, L., Dunlap, G., Hemmeter, M. L., Joseph, G. E., & Strain, P. S. (2003). The Teaching Pyramid: A model for supporting social competence and preventing challenging behavior in young children. *Young Children*, 58, 48-52.

Hemmeter, M. L., Fox, L., & Snyder, P. (2008). *Teaching Pyramid Tool for Preschool Classrooms (TPOT)*. Nashville, TN: Vanderbilt University.