

# Assuring High Quality in Publicly Funded Child Care and Preschool: A Cautionary Tale

Rachel A. Gordon

Professor, Department of Sociology

Associate Director, Institute of Govt and Public Affairs

University of Illinois at Chicago

*Presentation in Innovations in Design and Utilization of Measurement Systems to Promote Children's Cognitive, Affective and Behavioral Health an Institute of Medicine/National Research Council workshop (November 6, 2014).*



I N S T I T U T E O F  
GOVERNMENT & PUBLIC AFFAIRS

---

UNIVERSITY OF ILLINOIS  
URBANA-CHAMPAIGN • CHICAGO • SPRINGFIELD

# Recent Policy Context

- **Quality Rating and Improvement Systems (QRIS)** – which link child care subsidy levels to quality ratings – emerged in the late 1990s and now operate in about three-quarters of the states.
- The **Race to the Top Early Learning Challenge** encouraged states to integrate quality monitoring systems across funding streams, and encouraged all states to move toward QRIS.
- The **Improving Head Start for School Readiness Act of 2007** required lower quality Head Start grantees to recompete for funding (though none were actually required to until 2011).

# Leading to high stakes use of policy measures...

- The RTT-ELC required states to use “**valid and reliable**” indicators of the overall quality of the early learning environment and of the quality of adult-child interactions and that the validate whether the QRIS tiers reflect different levels of program quality and relate to **children’s progress in learning, development, and kindergarten readiness.**
- Head Start required “a **valid and reliable** research-based observational instrument... including assessing multiple dimensions of teacher-child interactions that are **linked to positive child development and later achievement.**”

# What is “reliable and valid”?

- The use of the term “reliable and valid” suggests that these are static properties of a measure for all time, all purposes, and all populations...

# What is “reliable and valid”?

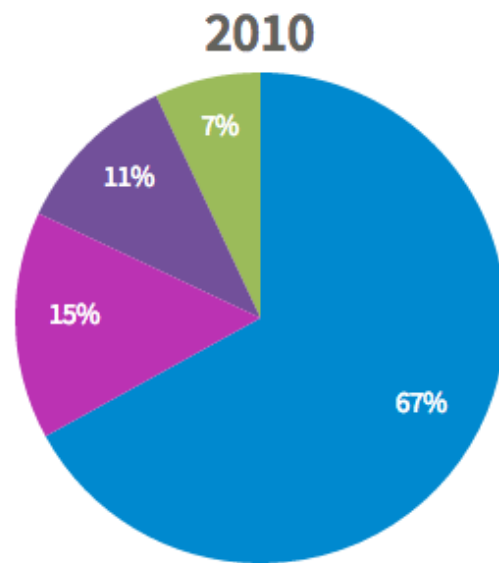
- Instead, consistent with the latest *Standards for Educational and Psychological Testing*, we should step back and consider:
  - the intents of each research and policy use
  - weigh the body of reliability and validity evidence against *each specific use*
  - build in continuous and local validation of measures selected for various uses
  - allow for the refinement of measures over place and time.

# In other words

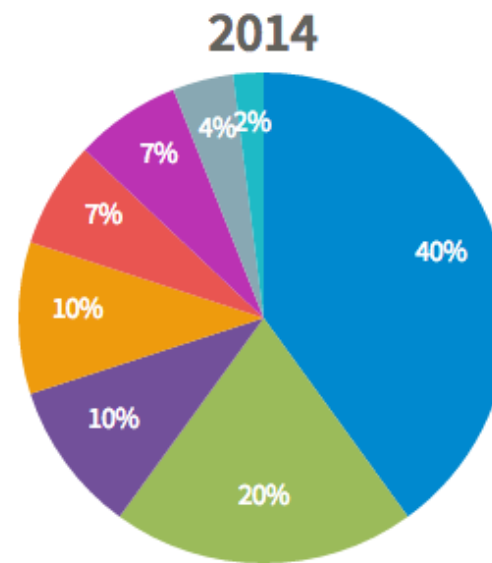
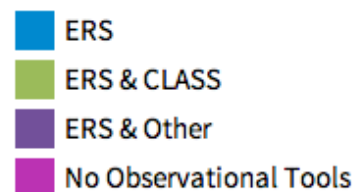
- The body of evidence desired to demonstrate reliability and validity for program self-assessment...
- May be different from reliability and validity for teacher professional development...
- Which may be different from reliability and validity for policy decision making and accountability...

# Focus on ECERS-R and CLASS

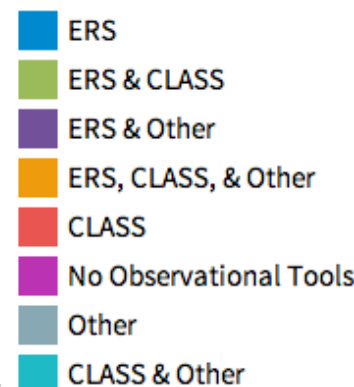
■ T  
p  
n  
n



Most commonly used observational tools in 2010.



Most commonly used observational tools in 2014.



art  
that

# What is the evidence for high stakes uses of the measures?

---

What do we know?  
What do we need to know?



# What we know

- I'm going to show you evidence indicating that...
  - In fact, the ECERS-R and CLASS are not highly associated with measures commonly used to assess school readiness gaps.
- The question then is: Why is this? I'll show evidence for some reasons related to potential limitations in the measures of quality:
  - Content of items.
  - Scoring procedures.
  - Inter-rater reliability.

# Preview

- The points I will make suggest that limitations of the reliability and validity evidence for current high stakes uses:
  - The ECERS-R may be covering the right content, but the standard scoring may not give centers credit for all of the features relevant to school readiness.
  - The CLASS may be focusing on important aspects of teacher-child interactions, but might benefit from more items in some areas and the inferential scoring may not be ideal for high stakes uses.
- I'll discuss how each measures origins for other uses may help us understand these limitations.

Small associations between  
ECERS-R/CLASS and  
standardized measures of  
child development.

# Evidence Base

- Earlier interpretations concluded stronger evidence of quality-outcome associations, but...
  - Often focused on statistical significance and not size of associations.
  - Often did not rigorously adjust for selection (confounds).
  - May have reflected better targeting of measures at typical quality several decades ago (prior to contemporary licensing and programmatic standards).
- Emerging consensus that contemporary quality-outcome associations are not always significant and generally small in size.
  - Often .10 or smaller in effect sizes (Abner et al., 2013; Burchinal, Kainz & Cai, 2011; Gordon et al., 2013; Keys et al., 2013).

# Limitations in quality measures:

Mixed content of items.

Scoring procedures.

Inter-rater reliability.

Skewness of items.

# Origins of the ECERS-R

- Developed in 1970s from a checklist to help practitioners improve the quality of their settings.
- Reflects *developmentally appropriate practice*, including:
  - predominance of child-initiated activities selected from a wide array of options;
  - a “whole child” approach that integrates physical, emotional, social and cognitive development.

# ECERS-R

- The organization of the ECERS-R items and its scoring procedures reflect its checklist, practice and philosophical origin.
  - There are over 400 indicators across 43 items.
  - These are grouped in ways that make sense to practice and philosophy, often organized around context of practice.
  - Helping to reduce burden, conditions in the indicators of lower scores must be met before indicators of higher scores are evaluated (thus not all indicators must be evaluated).

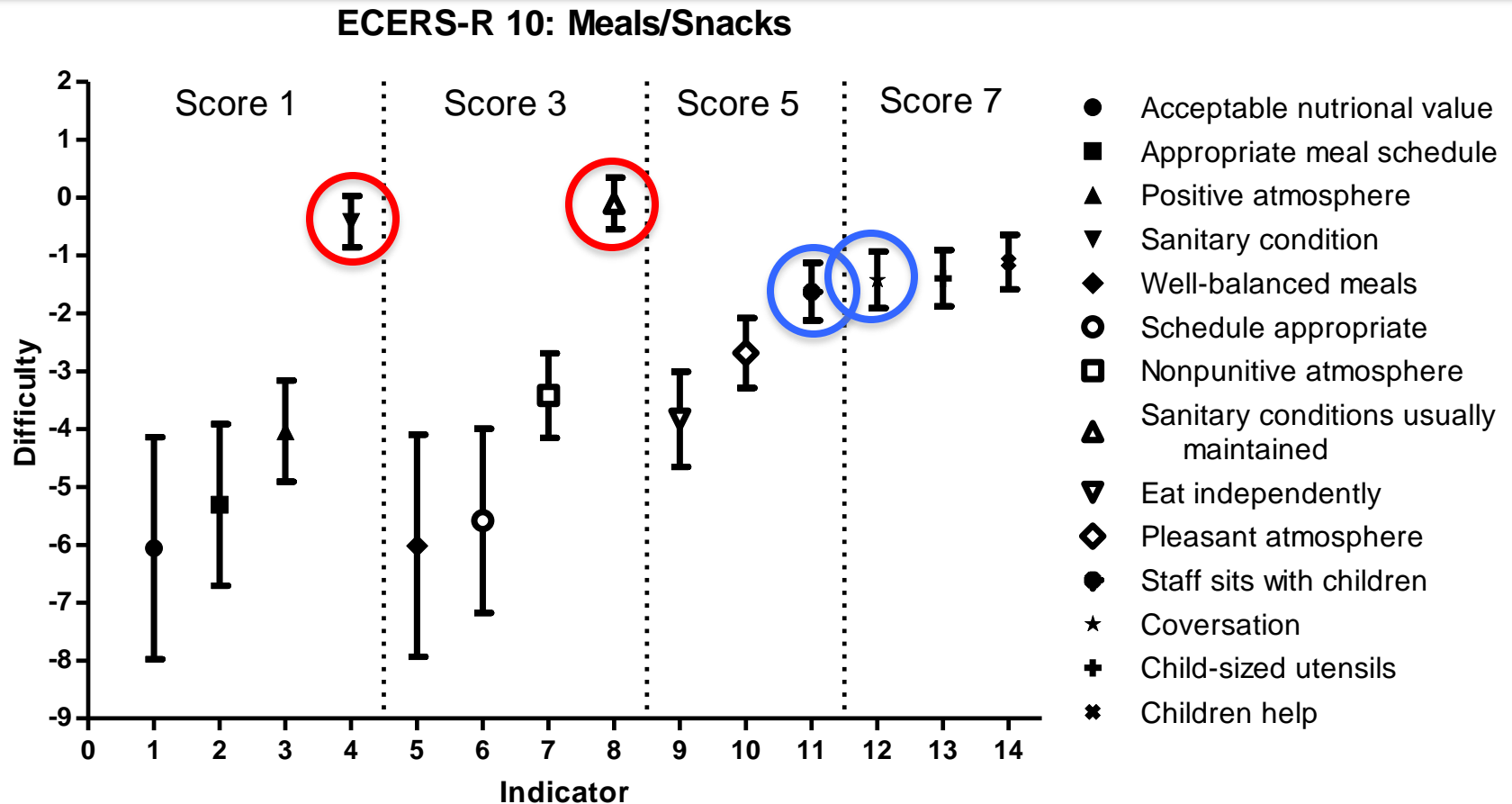
# ECERS-R Item 10: Meals/Snacks

Inadequate 1	2	Minimal 3	4	Good 5	6	Excellent 7
<b>10. Meals/snacks</b>						
1.1 Meal/snack schedule is inappropriate (Ex. child is made to wait even if hungry).		3.1 Schedule appropriate for children.	→	5.1 Most staff sit with children during meals and group snacks.‡	→	7.1 Children help during meals/snacks (Ex. set table, serve themselves, clear table, wipe up spills).
1.2 Food served is of unacceptable nutritional value.*		3.2 Well-balanced meals/snacks.*	→	5.2 Pleasant social atmosphere.		7.2 Child-sized <i>serving</i> utensils used by children to make self-help easier (Ex. children use small pitcher, sturdy serving bowls and spoons).
1.3 Sanitary conditions not usually maintained (Ex. most children and/or adults do not wash hands before handling food; tables not sanitized; toileting/diapering and food preparation areas not separated).		3.3 Sanitary conditions usually maintained.†		5.3 Children are encouraged to eat independently (Ex. child-sized <i>eating</i> utensils provided; special spoon or cup for child with disabilities).		7.3 Meals and snacks are times for conversation (Ex. staff encourage children to talk about events of day and talk about things children are interested in; children talk with one another).
1.4 Negative social atmosphere (Ex. staff enforce manners harshly; force child to eat; chaotic atmosphere).		3.4 Nonpunitive atmosphere during meals/snacks.		5.4 Dietary restrictions of families followed. <i>NA permitted.</i>		
1.5 No accommodations made for children's food allergies. <i>NA permitted.</i>		3.5 Allergies posted and food/beverage substitutions made. <i>NA permitted.</i>				
		3.6 Children with disabilities included at table with peers. <i>NA permitted.</i>				

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



# Evidence of Category Disorder



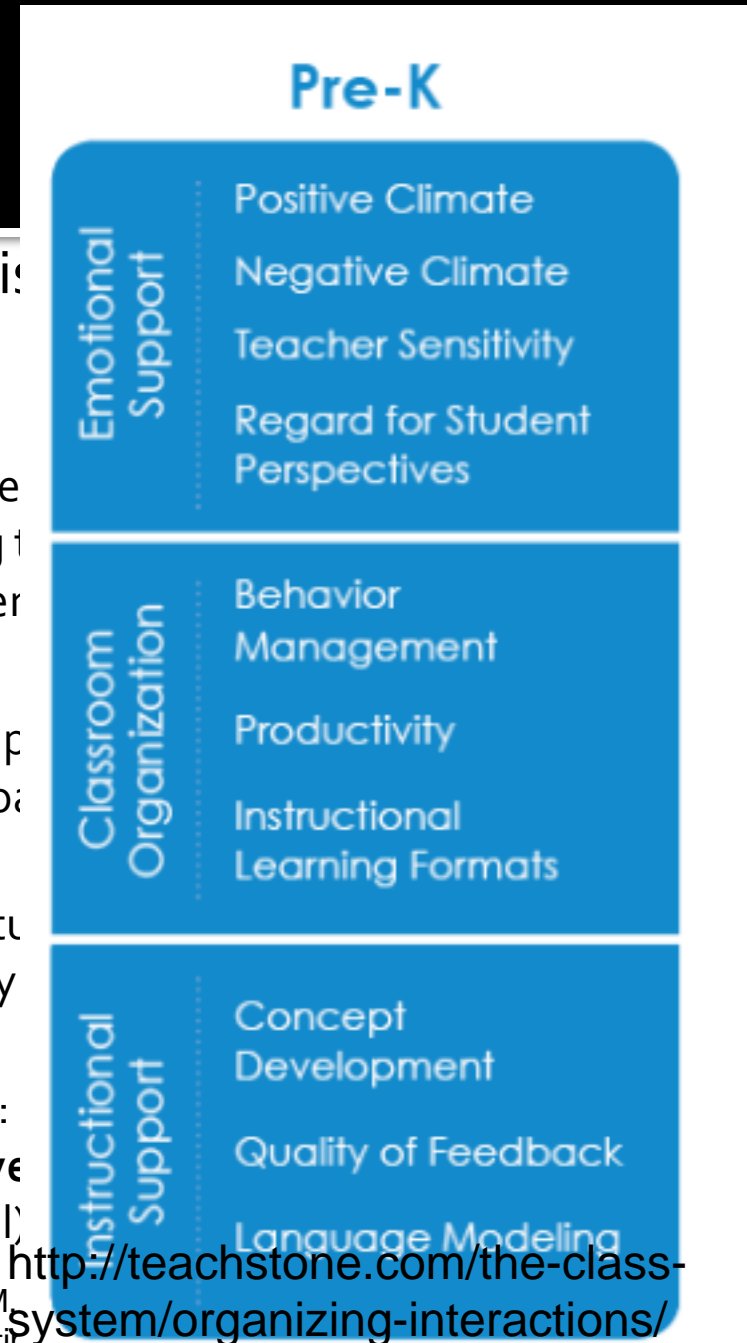
Source: Gordon, Rachel, Kerry Hofer, Ken Fujimoto, Nicole Colwell, Robert Kaestner, Sanders Korenman. "Measuring Aspects of Child Care Quality Specific to Domains of Child Development: An Indicator-level Analysis of the ECERS-R." Presented in the Paper Symposium "Measuring Early Care and Education Quality: New Insights about the Early Childhood Environment System Rating Scale - Revised" (Chair: Rachel Gordon Discussant: Margaret Burchinal) (Saturday April 20 2013, Seattle WA).

# Forthcoming ECERS-3

- The scale developers plan to release the ECERS-3 this fall.
- The revision may address some of these issues, although it is not clear yet how much of the ECERS-R structure is retained.
- And, the ECERS-R is written into current policy and embedded in many existing evaluations and studies.

# CLASS

- Unlike the checklist decades ago...
  - The CLASS was developed based on research suggesting that the classroom environment is a key mechanism of student learning.
  - Its predecessor was primarily focused on teacher development and coaching.
  - Very different structure to help observers assimilate what they see.
  - The manual advises: **should never be given a score** (p. 17, bold in original)



ECERS-R several

developmental theory and that children and adults are the primary focus (Pianta, La Paro & Hamre, p. 1)

was primarily aimed at professional development rather than policy contexts.

requires observers to rate many items.

**structure of the CLASS, scores** (Pianta, La Paro & Hamre,

# CLASS Results

- A recent publication from the CLASS developers (Cash, Hamre, Pianta, & Myers, 2012) reveals:
  - Exact reliability is low: 41% overall exact agreement with master score in training of 2,093 Head Start staff.
  - Black and Latino raters placed their scores farther from the master score as did raters who disagreed with intentional teaching beliefs.

# CLASS Results (cont).

- The CLASS developers also recently found (Hamre, Hatfield, Pianta & Jamil, 2014):
  - a bi-factor structure with one general dimension (responsive teaching) and two specific dimensions (proactive management and routines; cognitive facilitation).
  - these differ from the subscales written into policy
  - And domains may align differently than originally thought with aspects of quality specific to readiness.
- In our work, we are replicating these results.

# CLASS Results (cont.)

- We are also examining the targeting and content of items.

*Table 1.1 Score Distribution of Each CLASS Item for Categorical Data (Wave 21)*

Domain	Item	Within Scale Values							Total
		1	2	3	4	5	6	7	
Emotional Support	CLASS1: Positive Climate	0	1	5	87	217	92	6	408
	CLASS2: Negative Climate (R)	1	1	0	16	139	251	0	408
	CLASS3: Teacher Sensitivity	0	2	53	178	165	10	0	408
	CLASS4: Regard for Student Perspective	0	8	60	198	126	6	0	408
Classroom Organization	CLASS5: Behavioral Management	0	3	24	122	189	70	0	408
	CLASS6: Productivity	0	6	38	143	169	51	1	408
	CLASS7: Instructional Learning Format	3	39	110	165	67	4	0	408
Instructional Support	CLASS8: Concept Development	164	189	50	5	0	0	0	408
	CLASS9: Quality of Feedbacks	118	230	42	17	1	0	0	408
	CLASS10: Language Modeling	86	215	76	25	6	0	0	408

*Note.* (R) Indicates reverse scoring. Continuous scores were rounded down (retaining the integer, e.g., 6.90 becomes 6 and 7 represents all 7s).

# Circling back: What is “reliable and valid”?

- Consistent with the latest *Standards for Educational and Psychological Testing*, we should step back and consider:
  - the intents of each research and policy use
  - weigh the body of reliability and validity evidence against *each specific use*
  - build in continuous and local validation of measures selected for various uses
  - allow for the refinement of measures over place and time.



I N S T I T U T E O F  
GOVERNMENT & PUBLIC AFFAIRS

UNIVERSITY OF ILLINOIS  
URBANA-CHAMPAIGN • CHICAGO • SPRINGFIELD

WHERE LEADERS TURN FOR ANALYSIS OF CRUCIAL ISSUES

### The Early Investments Initiative

Connecting Evidence, Practice and Policy in Early Childhood Policy

- The AY14-15 project on the Chicago campus focuses on the quality of preschool and child care classrooms.
- The project has two specific components:
  - **Policy Brief:** independent and intensive examination of the criteria underlying several of the most common pathways in the state's quality rating system for centers/preschools and the ISBE state professional teaching standards in ECE.
  - **Pilot Study:** use of new technology to take a careful look at variation in quality within and across the school day and across quality definitions, measures and standards.



# Pre-K 'best practices' goal of PRI, MNPS team

by [Jennifer Johnston](#) | Posted on Friday, May. 16, 2014 — 4:28 PM



(iStockphoto)

Vanderbilt's [Peabody Research Institute](#) will collaborate with Metro Schools' newly appointed director of early learning innovation, Lisa Wiltshire, to create and document a preschool curriculum rooted in play, experimentation and discovery.

The expertise and prior experience of the researchers will allow quick analyses of child assessments and classroom observations to provide nearly immediate feedback to teachers.

<http://news.vanderbilt.edu/2014/05/pre-k-best-practices-team/>

Contact information:  
[ragordon@uic.edu](mailto:ragordon@uic.edu)



I N S T I T U T E O F  
GOVERNMENT & PUBLIC AFFAIRS

UNIVERSITY OF ILLINOIS  
URBANA-CHAMPAIGN • CHICAGO • SPRINGFIELD

# Acknowledgments

- Principal Investigators
  - Rachel Gordon
  - Kerry Hofer
- Other Investigators
  - Sandra Wilson
  - Everett Smith
- Graduate Students
  - Elisabeth Stewart
  - Jenny Kushto-Hoban
  - Hillary Rowe
  - Anna Colaner
- Graduate Students (con't).
  - Rowena Crabbe
  - Fang Peng
  - Danny Lambouths
  - Ken Fujimoto
- Consultant
  - Ariel Aloe
- **Institute for Education Sciences**  
**Grant #R305A130118**

# Acknowledgments

- Principal Investigator
  - Rachel Gordon
- Other Investigators
  - Everett Smith
  - Robert Kaestner
  - Sanders Korenman

Graduate Students  
Ken Fujimoto  
Kristin Abner  
Anna Colaner  
Nicole Colwell  
Xue Wang

**IES R305A090065**  
**NIH R01HD060711**

# Acknowledgments

- **Chicago Area Study**
- Principal Investigators
  - Rachel Gordon
  - Maria Krysan
- Project Director
  - Anna Colaner
- Project Team
  - Soc 501 and 509
  - Danny Lambouths
  - Denice Hutchings
- **Child Care Choices Pilot Study**
- Principal Investigators
  - Laura Stout Sosinsky
- Project Team
  - Rumeli Banik
  - Sonia Roubeni
  - Mergime Gjombalaj
  - Ruthanne Sobecki

**U.S DHHS, Administration for Children and Families**  
**ACF 90YE0144** Child Care Research Grant  
Dissertation Award for Anna Colaner

# Acknowledgments

- Principal Investigators
  - Roger Weissberg
  - Paul LaMarca
- Project Director:
  - Celene Domitrovich
- Consultants/co-Investigators
  - Rachel Gordon
  - Everett Smith
  - Rob Schamberg
  - Ann Bryson

Washoe County School District  
Ben Hayes  
Laura Davidson  
Trish Shaffer  
Marisa Crowder  
Randy Brown

**IES R305H130012**