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

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).
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.”

http://eclkc.ohs.acf.hhs.gov/hslc/standards/law
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

- The CLASS is also used to rate Head Start programs for the high stakes purpose of identifying those that must recompete for funding.

http://qriscompendium.org/top-ten/question-3/
What is the evidence for high stakes uses of the measures?

What do we know?
What do we need to 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.
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.
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**

<table>
<thead>
<tr>
<th>Inadequate</th>
<th>Minimal</th>
<th>Good</th>
<th>Excellent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

**10. Meals/snacks**

1. Meal/snack schedule is inappropriate (Ex. child is made to wait even if hungry).

2. Food served is of unacceptable nutritional value.*

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).

4. Negative social atmosphere (Ex. staff enforce manners harshly; force child to eat; chaotic atmosphere).

5. No accommodations made for children’s food allergies. **NA permitted.**

---

3.1 Schedule appropriate for children.

3.2 Well-balanced meals/snacks.*

3.3 Sanitary conditions usually maintained†

3.4 Nonpunitive atmosphere during meals/snacks.

3.5 Allergies posted and food/beverage substitutions made. **NA permitted.**

3.6 Children with disabilities included at table with peers. **NA permitted.**

---

5.1 Most staff sit with children during meals and group snacks.‡

5.2 Pleasant social atmosphere.

5.3 Children are encouraged to eat independently (Ex. child-sized eating utensils provided; special spoon or cup for child with disabilities).

5.4 Dietary restrictions of families followed. **NA permitted.**

---

7.1 Children help during meals/snacks (Ex. set table, serve themselves, clear table, wipe up spills).

7.2 Child-sized serving utensils used by children to make self-help easier (Ex. children use small pitcher, sturdy serving bowls and spoons).

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).

---

Evidence of Category Disorder

ECERS-R 10: Meals/Snacks

Score 1
Score 3
Score 5
Score 7

Acceptable nutritional value
Appropriate meal schedule
Positive atmosphere
Sanitary condition
Well-balanced meals
Schedule appropriate meals
Nonpunitive atmosphere
Sanitary conditions usually maintained
Eat independently
Pleasant atmosphere
Staff sits with children
Conversation
Child-sized utensils
Children help

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.
Unlike the checklist and practice origins of the ECERS-R several decades ago...

- The CLASS was developed more recently based on "developmental theory and research suggesting that interactions between students and adults are the primary mechanism of student development and learning." (Pianta, La Paro & Hamre, p. 1)

- Its predecessor was part of a research study, and it was later aimed at professional development and coaching before being adopted in high stakes policy contexts.

- Very different structure than ECERS-R: the CLASS manual requires observers to assimilate what they see in order to assign scores to just a few items.

- The manual advises: "Because of the highly inferential nature of the CLASS, scores should never be given without referring to the manual." (Pianta, La Paro & Hamre, p. 17, bold in original)


http://teachstone.com/the-class-system/organizing-interactions/
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.
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.
We are also examining the targeting and content of items.

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

<table>
<thead>
<tr>
<th>Domain</th>
<th>Item</th>
<th>Within Scale Values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Emotional Support</td>
<td>CLASS1: Positive Climate</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CLASS2: Negative Climate (R)</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>CLASS3: Teacher Sensitivity</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CLASS4: Regard for Student Perspective</td>
<td>0</td>
</tr>
<tr>
<td>Classroom Organization</td>
<td>CLASS5: Behavioral Management</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CLASS6: Productivity</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>CLASS7: Instructional Learning Format</td>
<td>3</td>
</tr>
<tr>
<td>Instructional Support</td>
<td>CLASS8: Concept Development</td>
<td>164</td>
</tr>
<tr>
<td></td>
<td>CLASS9: Quality of Feedbacks</td>
<td>118</td>
</tr>
<tr>
<td></td>
<td>CLASS10: Language Modeling</td>
<td>86</td>
</tr>
</tbody>
</table>

Note. (R) Indicates reverse scoring. Continuous scores were found 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.

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

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
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