What is Program Quality and How Do We Measure It?

Rachel A. Gordon
Professor, Department of Sociology
Associate Director, Institute of Govt and Public Affairs
University of Illinois at Chicago

Presentation in the Research Symposium on “Research on Program Quality and Program Standards” at the NAEYC National Institute for Early Childhood Professional Development (June 10, 2014).

INSTITUTE OF GOVERNMENT & PUBLIC AFFAIRS
UNIVERSITY OF ILLINOIS
URBANA-CHAMPAIGN • CHICAGO • SPRINGFIELD

The Importance of Program Quality

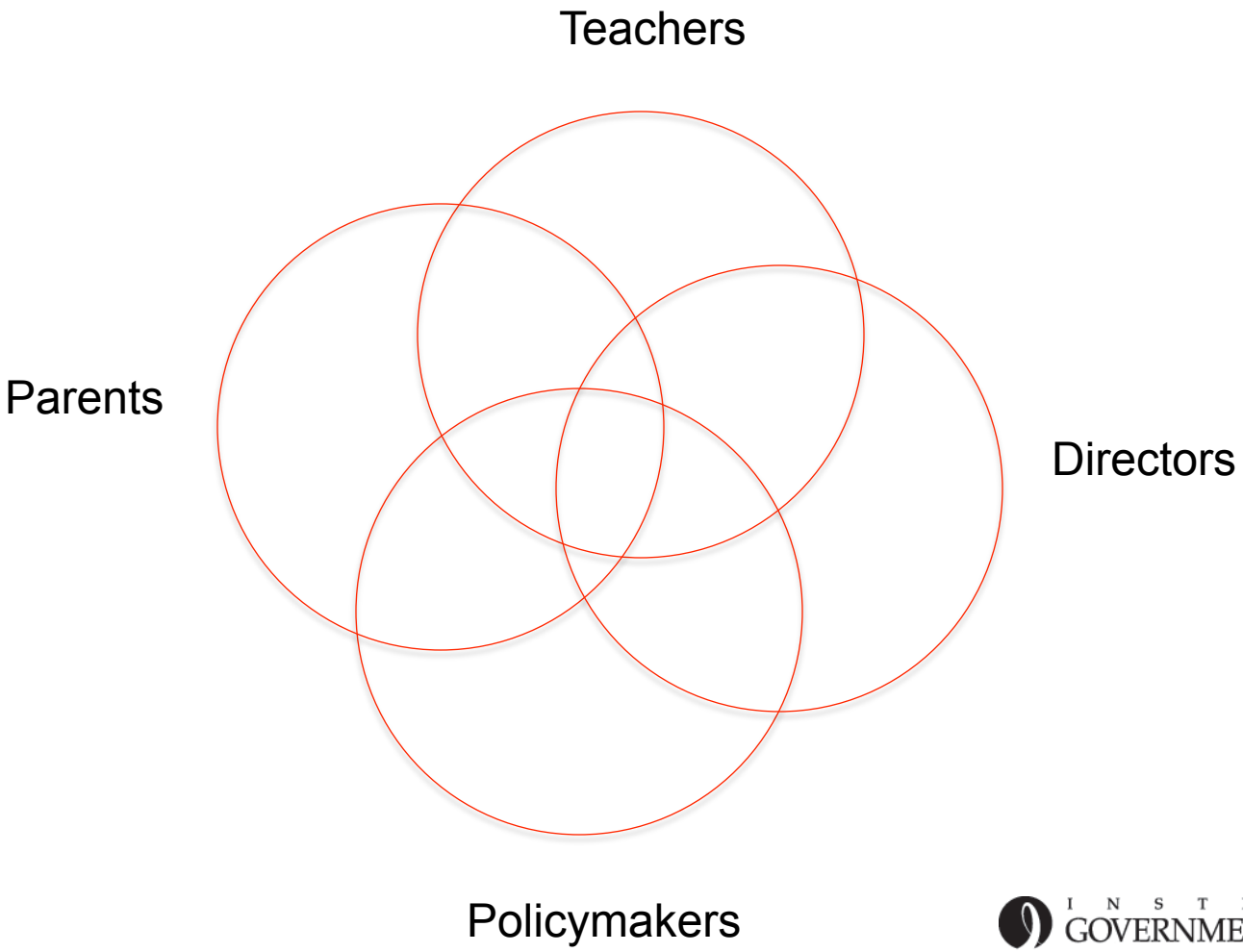
- Teachers, Directors, Parents, Policymakers
  - All share a desire for high quality early care and education.
- But what does each stakeholder mean by high quality?
- And how to do we know if programs are of high quality now, or are moving toward it?

Institute of Government & Public Affairs
University of Illinois
Urbana-Champaign • Chicago • Springfield

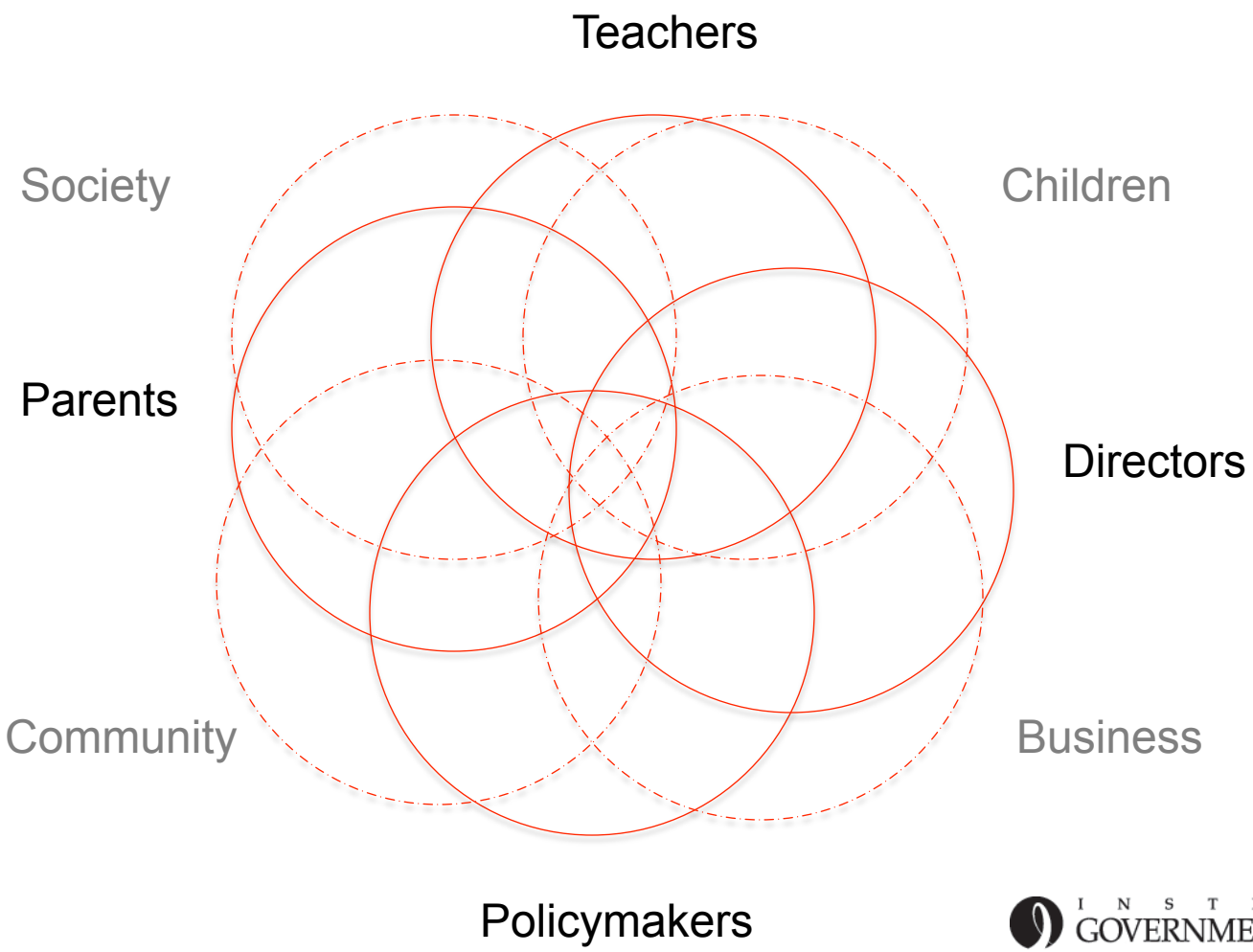
The Big Picture

- Of course quality sits at the center of NAEYC's origins, as the organization's earliest work set minimum standards for nursery education.
- The contemporary NAEYC standards and accreditation system, and position statements including on developmentally appropriate practice, continue that work.
- These NAEYC efforts intersect with many other conceptions of quality.

The Big Picture: Stakeholders

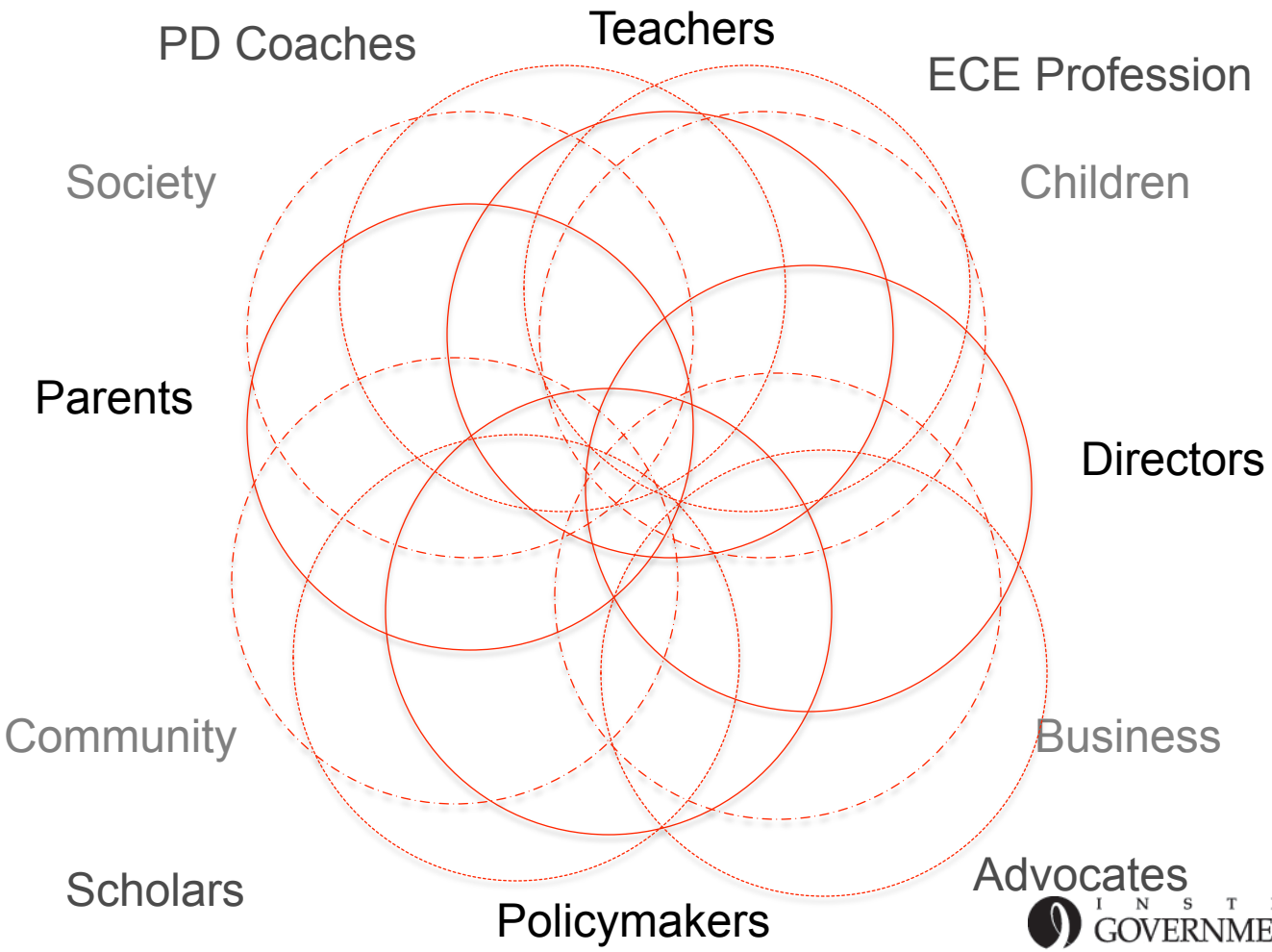


The Big Picture: Stakeholders



INSTITUTE OF GOVERNMENT & PUBLIC AFFAIRS
UNIVERSITY OF ILLINOIS
URBANA-CHAMPAIGN • CHICAGO • SPRINGFIELD

The Big Picture: Stakeholders



INSTITUTE OF GOVERNMENT & PUBLIC AFFAIRS
UNIVERSITY OF ILLINOIS
URBANA-CHAMPAIGN • CHICAGO • SPRINGFIELD

Big Picture: Three Broad Pieces of Quality f ren

Language
Literacy
Math
Approaches to Learning
Self-Regulation
Health

- **Optimal Development** sets the stage to suppo and curiosity.
- **School Readiness:** Perhaps ned to help children acquire the skills ded at school entry, often with particular attention to closing readiness gaps.
- **Work Support:** Such as a clean and safe environment where children are happy and healthy ("have a good day") while their parents work

INSTITUTE OF 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 half 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.**"

https://www.federalregister.gov/articles/2011/08/26/2011-21756/applications-for-new-awards-race-to-the-top-early-learning-challenge
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 these uses
  - allow for the refinement of measures over place and time.

http://www.apa.org/science/programs/testing/standards.aspx

In other words

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

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 some evidence for some reasons. Some possibilities include:
  - Are the measures of school readiness limited?
  - Are the measures of quality limited?
    - Content of items.
    - Scoring procedures.
    - Inter-rater reliability.
    - Variation across classrooms in centers.
    - Variation within classrooms.

Preview

- The points I will make suggest that:
  - 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.

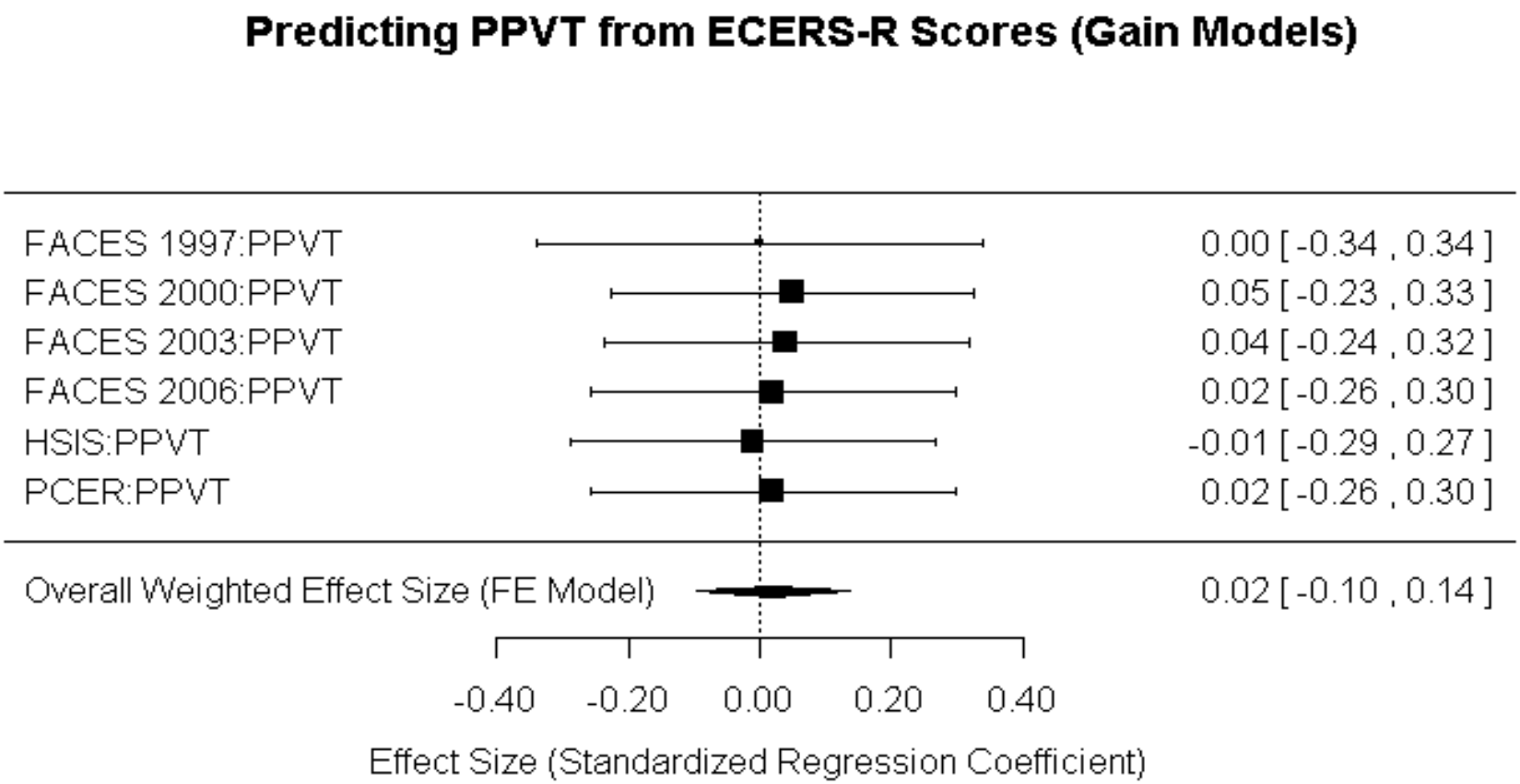
Small associations between ECERS-R/CLASS and standardized measures of child development.

Prior Studies

- Burchinal, Kainz and Cai (2011)
  - Effect sizes (adjusted correlations) of .14 and below across published studies correlating various quality measures with child outcomes.
  - Even when focusing on low-income children and aligning subscales with language, math, social, and behavioral outcomes in new analyses, 32 of 36 adjusted correlations at or below .10.

- Keys and colleagues (2013)
  - Average effect sizes between .01 to .05 for language, math, social, and behavioral outcomes.

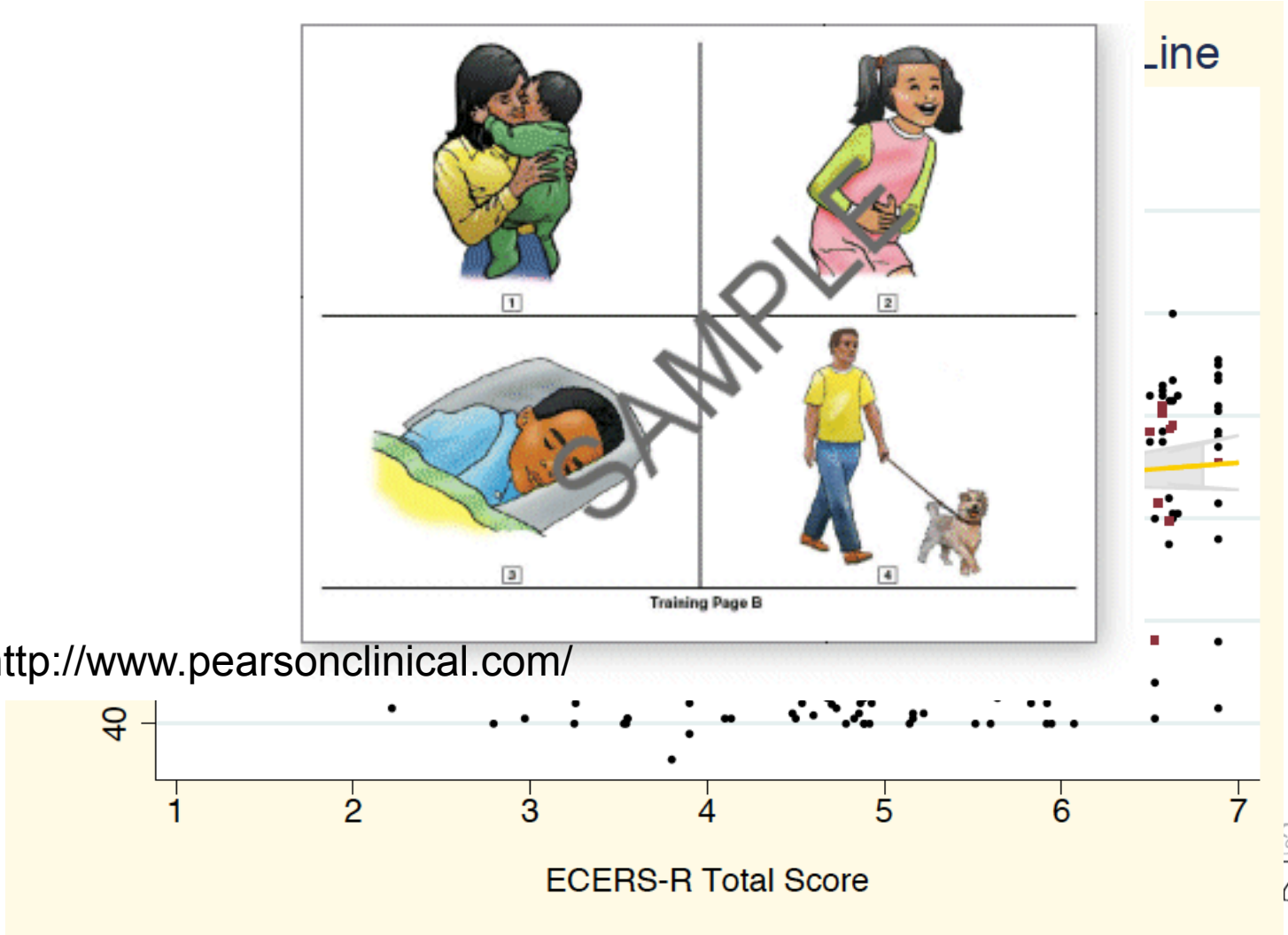
Our new results



Small Associations In All Regions of Quality

PPVT-4 Sample Training Item

http://www.pearsonclinical.com/



UTE OF PUBLIC AFFAIRS
Y OF ILLINOIS
CHICAGO • SPRINGFIELD

Limitations in quality measures:

Mixed content of items.

Scoring procedures.

Inter-rater reliability.

Origins of the ECERS-R

- Developed in 1970s from a checklist to help practitioners improve the quality of their settings.

- Reflects *developmentally appropriate practice*:
  - predominance of child-initiated activities selected from a wide array of options;
  - a “whole child” approach that integrates physical, emotional, social and cognitive development;
  - teacher facilitation of development by being responsive to

INSTITUTE OF GOVERNMENT & PUBLIC AFFAIRS
UNIVERSITY OF ILLINOIS
URBANA-CHAMPAIGN | CHICAGO | SPRINGFIELD

ECERS-R

- Standard "stop scoring" struc
this checklist, practice and ph
origin.

Over 400 indicators across the 43 items!

  - Categories from 1 to 7 have several "indicators"

  - Conditions in the indicators of lower scores must be met before indicators of higher scores are evaluated.

- Especially within some items, indicators often organized around contexts of practice and reflect multiple aspects of quality.

INSTITUTE OF GOVERNMENT & PUBLIC AFFAIRS
UNIVERSITY OF ILLINOIS
URBANA-CHAMPAIGN · CHICAGO · SPRINGFIELD

ECERS-R Item 10: Meals/ Snacks

Inadequate 1	2	Minimal 3	4	Good 5	6	Excellent 7

10. Meals/snacks

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

1.2 Food served is of unacceptable nutritional value.*

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).

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

1.5 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).

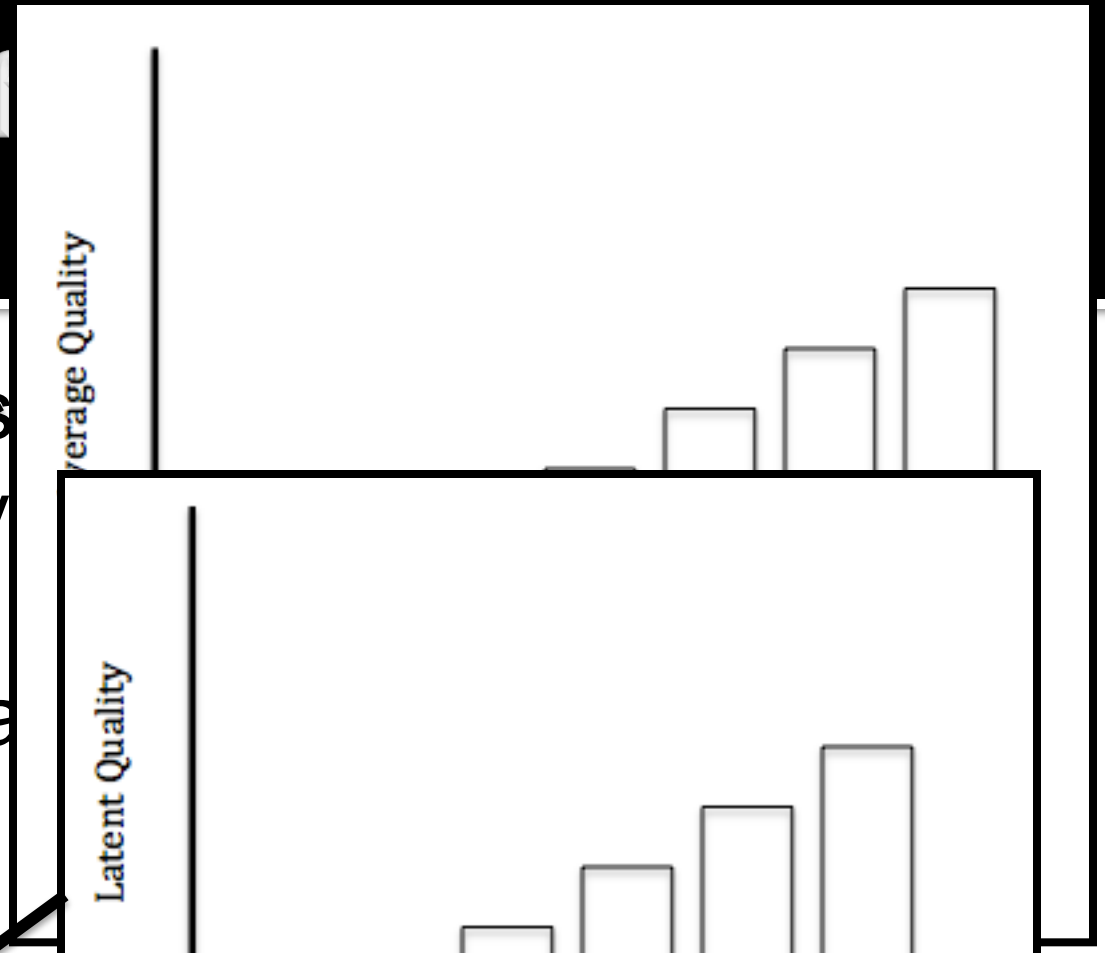
7.3 Meals and snacks are times for conversation (Ex. staff encourage children to talk about events of day and talk about things children are interested in; children talk with one another).

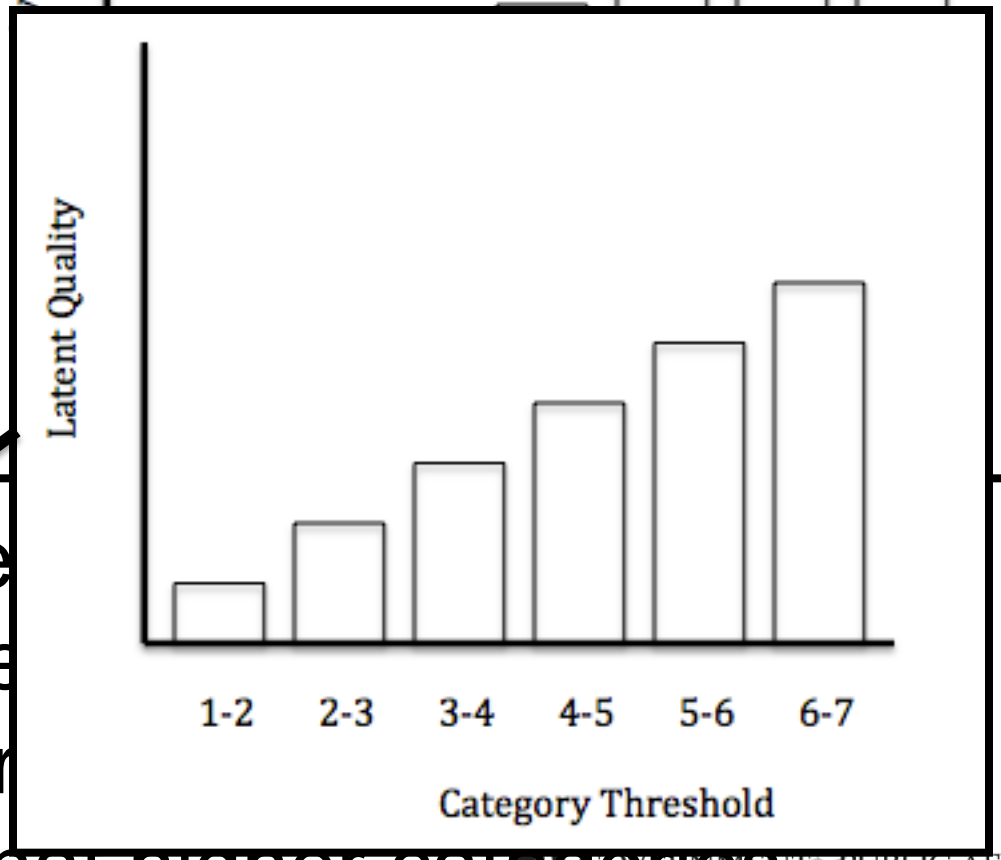
Source: Harms, T., Clifford, R.M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale, Revised Edition.* New York, NY: Teachers College Press.

Evidence of Disorder

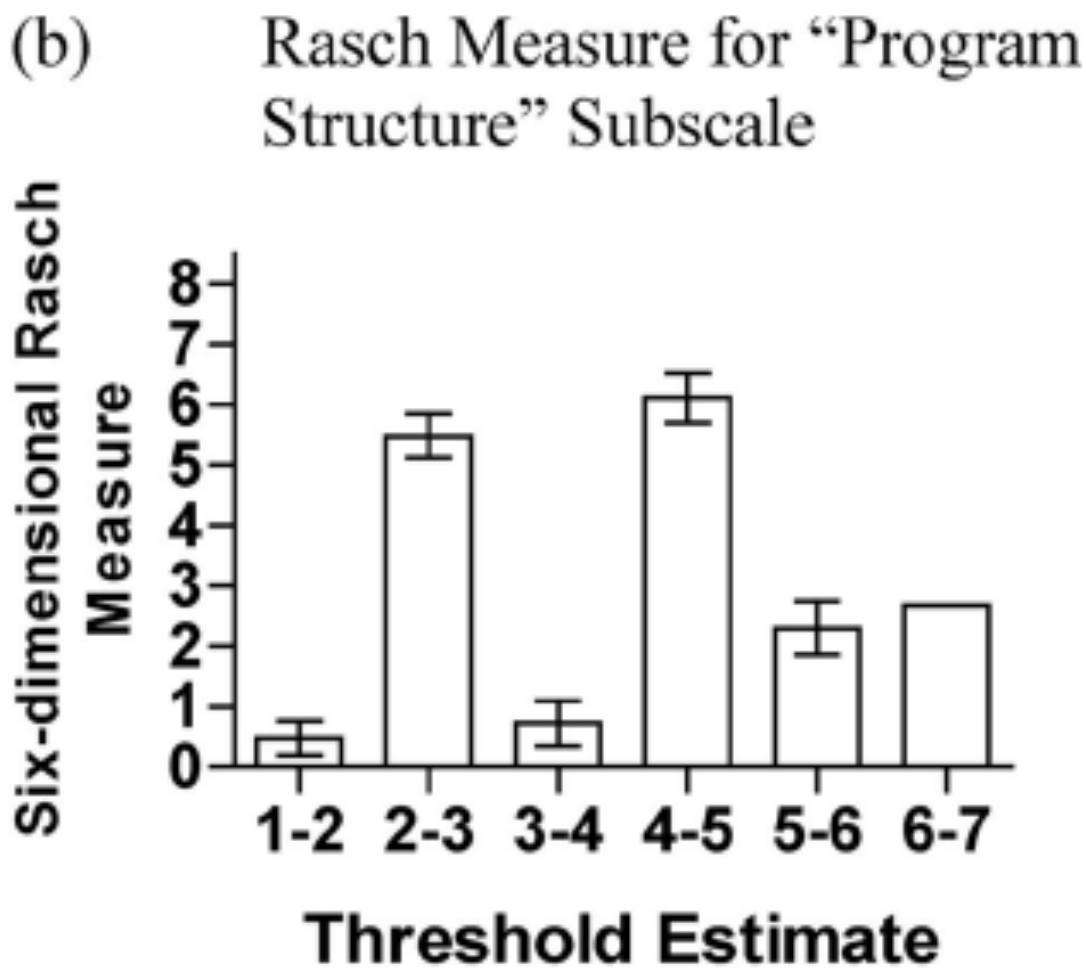
- If higher scores average quality among centers versus lower ca item.

- In item response thresholds betwe also show a stair are ordered so that higher categories mark higher quality





Evidence of Category Disorder

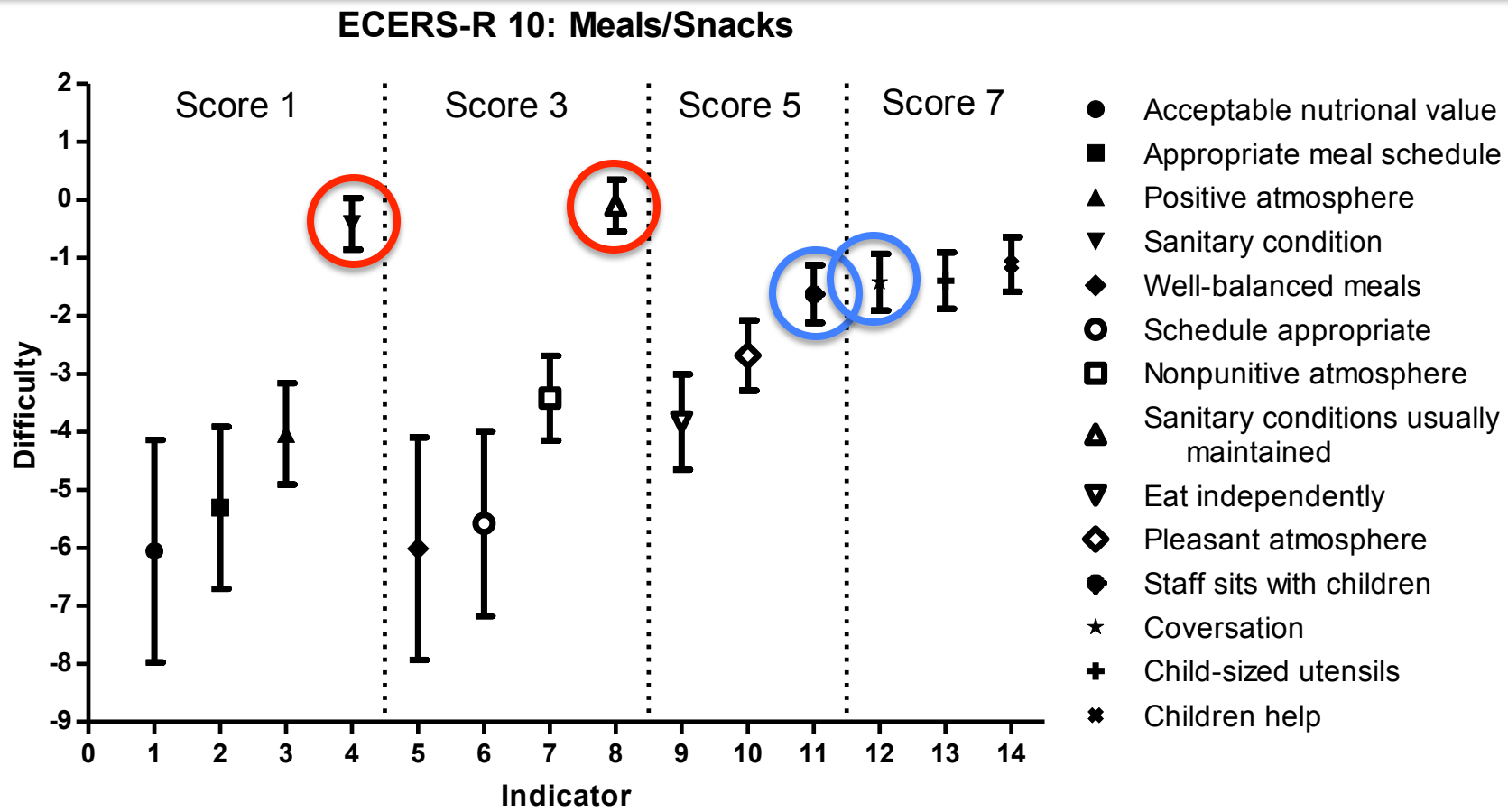


Source: Gordon, Rachel
2013. "An Assessment of
Care Quality and its Relatic

Replication of Category Disorder

Study Name	Initial Year	Focal Population
3-City Study	1999	Low-income families from low-income neighborhoods in Boston, Chicago and San Antonio.
ECLS-B	2001	Nationally-representative sample drawn from birth records in 46 states.
Early Head Start Research Eval Project	1996-1998	New Early Head Start applicants with a child under 12 months of age.
FACES 1997	1997	New Head Start 3- and 4- year old participants.
FACES 2000	2000	
FACES 2003	2003	
Head Start Impact Study	2002	
Fragile Families	1998-2000	Birth records sampled from hospitals in twenty large U.S. cities.
PCER	2003	Twelve sites implemented curricula in preschool programs. Each site had 14-20 programs.
QUINCE	2004	Twenty-four CCR&R agencies in five states (CA,IA,MN,NE,NC)

Evidence of Category Disorder



SourceGordon, 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).

CLASS

- 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 aimed at professional development and coaching use before being adopted in high stakes policy contexts.
  - 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)

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

Institute of Government & Public Affairs
University of Illinois
Urbana-Champaign • Chicago • Springfield

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 Instructional Support scores farther from the master score as did raters who disagreed with intentional teaching

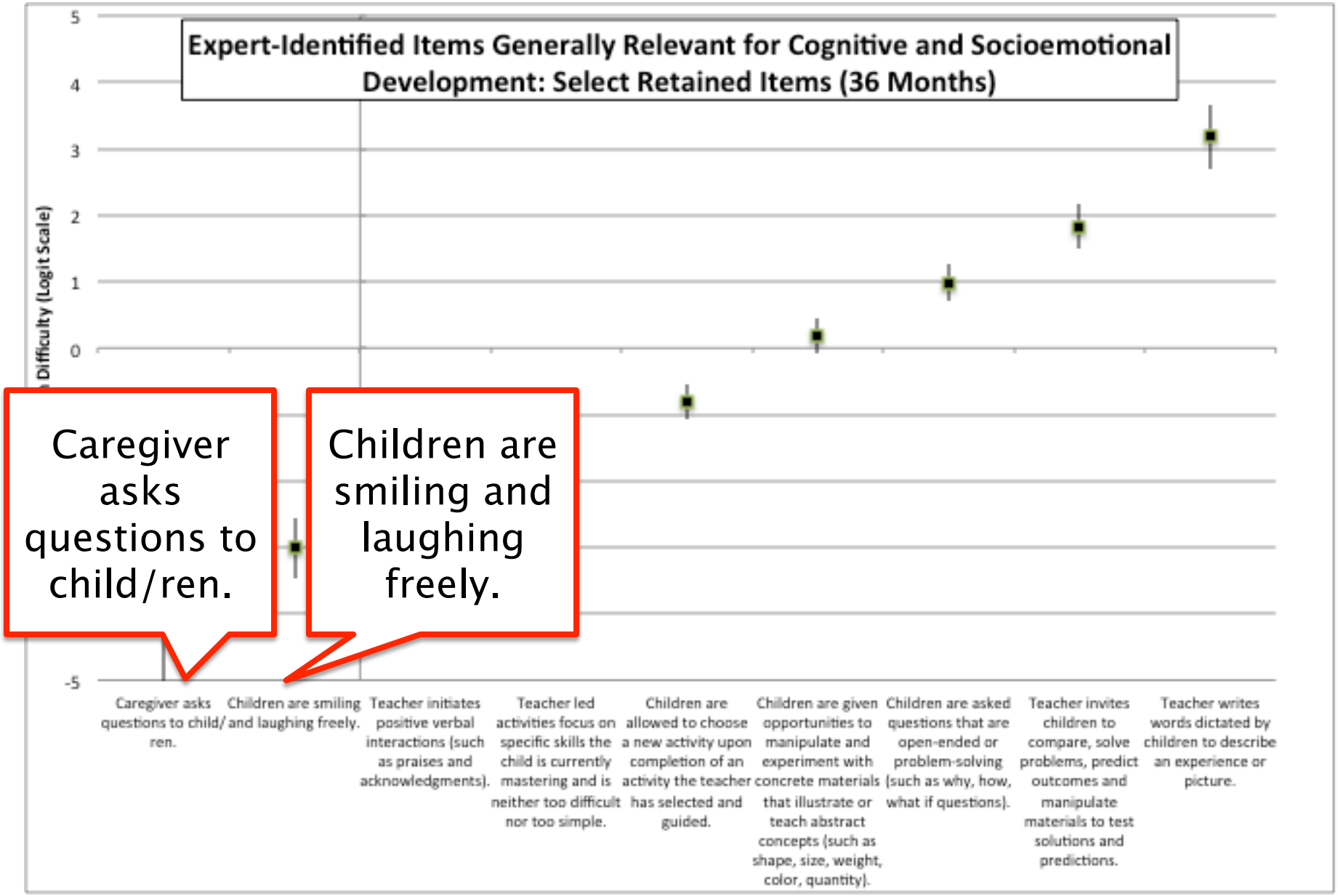
INSTITUTE OF GOVERNMENT & PUBLIC AFFAIRS
UNIVERSITY OF ILLINOIS
URBANA-CHAMPAIGN · CHICAGO · SPRINGFIELD

CLASS Results (cont).

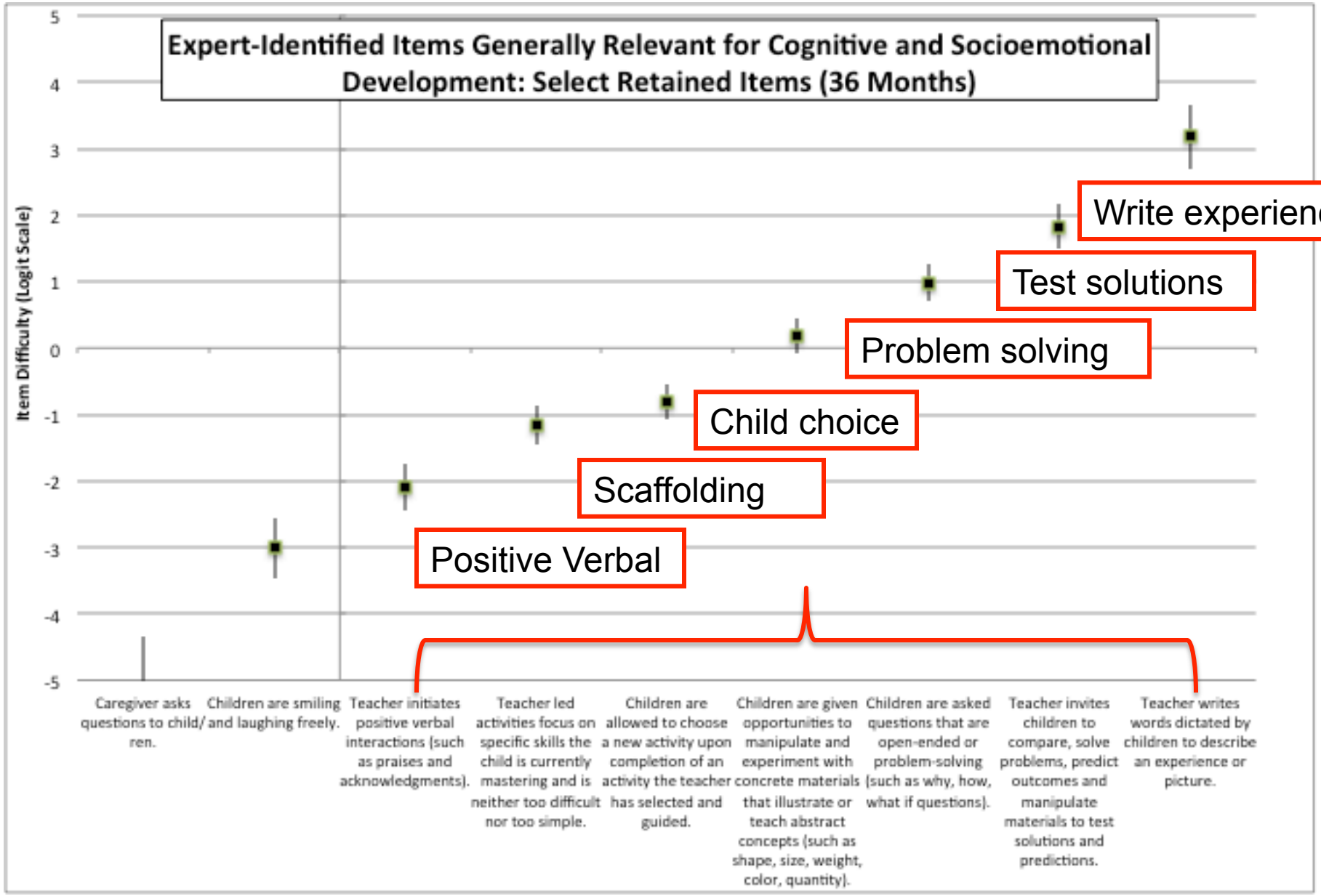
- The CLASS developers also recently found (Hamre, Hatfield, Pianta & Jamil, 2014):
  - a bi-factor structure with one general dimension (responsive teaching) and two specific dimensions (proactive management and routines; cognitive facilitation).
  - these differ from the subscales written into policy.
- In our work, we are replicating these results, and also examining the targeting and content of items with IRT models.

Next steps: What we need to know

- Even with the current limitations and challenges, there is much potential.
  - Continued building of evidence and insights into existing measures.
  - Refining definitions of relevant aspects of quality.
  - Developing of new measures.



Source: Colwell, Nicole, Rachel Gordon, Ken Fujimoto, Robert Kaestner, and Sanders Korenman. "Domain-Specific Quality Measures for Early Childhood Programs: New Evidence from the Study of Early Child Care and Youth Development." Presented in the Paper Symposium "New Insights into Early Care and Education Quality and Child Development: Profiles of Care and Domain-Specific Aspects of Quality" (Chair: Rachel A. Gordon, Discussant: Timothy W. Curby). (Thursday, April 18, 2013, Seattle WA).



Source: Colwell, Nicole, Rachel Gordon, Ken Fujimoto, Robert Kaestner, and Sanders Korenman. "Domain-Specific Quality Measures for Early Childhood Programs: New Evidence from the Study of Early Child Care and Youth Development." Presented in the Paper Symposium "New Insights into Early Care and Education Quality and Child Development: Profiles of Care and Domain-Specific Aspects of Quality" (Chair: Rachel A. Gordon, Discussant: Timothy W. Curby). (Thursday, April 18, 2013, Seattle WA).

Contact information:
ragordon@uic.edu

INSTITUTE OF GOVERNMENT & PUBLIC AFFAIRS
UNIVERSITY OF ILLINOIS
URBANA-CHAMPAIGN • CHICAGO • SPRINGFIELD

Acknowledgments

The opinions expressed are those of the authors and do not represent views of the Institute, the U.S. Department of Education, or NIH.

- 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
  - Betsy Becker

- **Institute for Education Sciences Grant #R305A130118**

INSTITUTE OF GOVERNMENT & PUBLIC AFFAIRS
UNIVERSITY OF ILLINOIS
URBANA-CHAMPAIGN • CHICAGO • SPRINGFIELD

Acknowledgments

The opinions expressed are those of the authors and do not represent views of the Institute, the U.S. Department of Education, or NIH.

- 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

INSTITUTE OF GOVERNMENT & PUBLIC AFFAIRS
UNIVERSITY OF ILLINOIS
URBANA-CHAMPAIGN • CHICAGO • SPRINGFIELD

Acknowledgments

The opinions expressed are those of the authors and do not represent views of the study funders.

- **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

INSTITUTE OF GOVERNMENT & PUBLIC AFFAIRS
UNIVERSITY OF ILLINOIS
URBANA-CHAMPAIGN • CHICAGO • SPRINGFIELD

Acknowledgments

The opinions expressed are those of the authors and do not represent views of the Institute, the U.S. Department of Education, or our consultants

- 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

INSTITUTE OF GOVERNMENT & PUBLIC AFFAIRS
UNIVERSITY OF ILLINOIS
URBANA-CHAMPAIGN • CHICAGO • SPRINGFIELD