Welcome!

Measuring Quality in QRIS Contexts

CA-QRIS
December 20, 2016
12:30pm-2:30pm

Participant Introductions

Please use the chat box to introduce yourself.

1. Look for the chat box at the bottom right corner of your screen.

2. Type the following into the box:

   o Name

   o Role

   o Region or organization

   o Workgroup (name of workgroup of which you are a member)

3. Please make sure you select "Everyone" when you send.

Introductions

Sarah Neville-Morgan, First 5 California

Guest Speaker:

Diana Schaack, University of Colorado

MEASURING QUALITY IN QRIS CONTEXTS

Diana Schaack, PhD

University of Colorado Denver

Purpose

To consider:

- How evaluation can be used to support states in QRIS revision

- What other states have learned about how to effectively measure quality within a QRIS

- Additional areas where research is needed

Looking forward:

- How California can leverage existing data to support QRIS revision

- What California may want to learn about measuring quality before making revisions

Agenda

- Framework for revising ratings

- Overview of RTT-ELC AIR evaluation

- Measuring quality in QRIS contexts
  - Individual rating components
  - Setting cut-points on quality measures

- Key priorities for discussions for workgroup meeting

Revising rating approaches

Your values:

☐ Supports closing the opportunity and achievement gaps, particularly for Latino and African American children

☐ Promotes a culture of quality and equity

☐ Promotes developmentally appropriate practice, such as play-based learning

☐ Promotes culturally and linguistically appropriate practices

☐ Connects to the current early learning and development system

☐ Promotes inclusion

Framework for revision

CA-QRIS Matrix Workgroup Guiding Principles	Revision decision points (Schaack & Tarrant, 2014)
Is it linked to improve outcomes: child, family, system	Theoretical links to QRIS goals/desired outcomes
Is it research based Is it measureable along a continuum	Evaluation results
Do benefits outweigh cost	Cost
Is it scalable and sustainable	Feasibility
Is it inclusive	Culturally Competent
	Perceptions of System Beneficiaries/Users
	Regulatory Context
	Special Interest

Today's Focus

Using research and evaluation to guide considerations about QRIS revisions

The Quality Rating

QRIS is founded on the assumption that:

□ Quality of early childhood programs can be measured

□ Results from the measurements are credible

□ Differences between programs that are measured through the QRIS are accurate and associated with meaningful outcomes.

Purpose of QRIS validation
(adapted from Tout et al., 2015)

QRIS validation should be considered an ongoing and iterative process that provide states with information about whether QRIS design decisions are producing meaningful changes……

How can we use evaluation results?

Step 1:

- ☐ Content validity

Step 2:

- ☐ Properties of measures

- ☐ Rating distinct aspects of quality

- ☐ Meaningful distribution of ratings

- ☐ Quality indicators related to previously validated measures of quality

- ☐ Quality indicators generally related to child outcomes

Where do we start?

Making sure the components work well! The case of the ECERS-R (Hoffer, 2010)

Data collection matters!

❑ Characteristics of observation
- ❑ How long raters observed and when they started observation
- ❑ Start time unrelated to overall ECERS-R score
- ❑ Longer observations related to lower-ECERS-R scores
- ❑ Impacted overall quality rating and center funding

Another Example

- Score up versus stop scoring method
  - Great deal of variability in whether raters followed the stop rule (Gordon et al., 2013)
  - The score-up approach yielded higher ECERS scores than stop scoring;
  - 28% of programs moved to a different rating as a result of using different scoring rules (Hoffer, 2010)

Implications: Scoring and administration consistency across communities is critical. Lack of consistency means that a quality rating in one community may not be the same as in another.

Types of Measures

- Self-reports
  - Often little concordance between self-assessment and independent observer scores, with self-assessment scores higher (Faria, et al., 2015)

Let's Consider Classroom Ratios

- No relationships between ratios and observed classroom quality and child outcomes using self-report (Zellman et al., 2008)

- QRIS evaluation as an iterative process:
    - Smaller Pilot Study of time sampling
    - Self-report under-estimated ratios in comparison to one-time direct observations
    - One-time direct observations underestimated in comparison to multiple time samples

- Time sampling approach found relationships with classroom quality (Le, et al., 2006)

Revision Process

- Considered evaluation and evaluators as tools
- Did not initially change rating criteria for ratios when no relationships to outcomes were found
- Asked additional research questions; considered what additional data would be needed
- Conducted a pilot
- Revised ratio rating approach

Training and Education: A tricky nut

Training and Education

☐ Self-reports not always accurate

☐ States that link QRIS with professional development systems have transcripts evaluated

☐ States often take a more is better approach in rating amount of formal education

   ☐ Great deal of variation in content of degrees (Whitebook & Ryan, 2011)

Training and Education

Colorado example:

- No relationships between teacher education and child outcomes (Zellman, et al., 2008)

- Time Sampling: Movement of teachers and children in and out of different classrooms

- Once the dosage of time teachers spent with children was considered, relationships between teacher education levels, ERS and child outcomes were found (Setodji, Le, & Schaack, 2012)

Implications: Should continuity of care be added as a dimension of quality?

Important Questions Remain

- What types of educational content drives better teaching and learning outcomes?

- How do we weight different positions (e.g., teachers, directors, assistant teachers) in center-based settings?
  - Some states consider head teachers only
  - Some states have higher standards for head teachers
  - Some states consider head and assistant teachers the same

Cut Scores

Cut-points

- Major aspect of QRIS theory of change
  - Cut scores determine an overall rating
  - Cut scores determine component level points

- Theoretically cut-points imply a quantitatively different level of quality
  - For example, we assume that children in star 2 programs will outperform children in star 1 programs because they have experienced qualitatively better quality ECE

Quality Ratings

- ☐ Most states have not found relationships between overall quality ratings and child outcomes (Karoly, 2014)

- ☐ Some have found differences in child outcomes between lowest and highest quality rated programs (e.g., North Carolina)

- ☐ May be important to consider examining individual component cut-scores

Examining Cut-scores

- QRIS analytic assumption: more is better
  - No significant linear relationships between teacher training and education, classroom quality, and child outcomes (Zellman, et al, 2008)

- Looking at thresholds:
  - When teachers took 1-9 ECE credits, we see an increase in ECERS-R scores, but after 9 credits, no relationships to ECERS-R scores (Setodji, Le, & Schaack, 2013)
  - Once teachers took more than 15 ECE credits we observed increases in child outcomes (preliminary analysis)

ECERS-R Thresholds

- Colorado QRIS: No significant linear relationships between ECERS-R and child outcomes (Zellman et al, 2008)

- Looking at thresholds:
  - Colorado: Once ECERS-R scores reached 3.40 and 5.40 related to better cognitive development (Le, Schaack, & Setodji, 2013)
  - National Sample: Structural quality subscales tended to level off around 3.40; process oriented subscales had higher thresholds related to better child outcomes, around 5.00 (Le, Schaack, Setodji, in press)

CLASS Thresholds

Thresholds for instructional quality and emotional support:

❑ Instructional quality was related to expressive language, reading, and math skills once scores exceeded a 3.00

❑ Emotional support positively predicted social competence and negatively predicted behavioral problems at a score of 5.00

Implications across threshold studies: There may be fewer quality cut-points related to child outcomes than reflected in most states' QRIS

❑ Small changes in cut-scores can mean a sizable portion of programs qualify for a different star rating (Fiara et al., 2015)

Important Questions

□ How do we combine the CLASS and ECERS-R?

State approaches:

□ After a certain ECERS-R score, the CLASS is administered

- ◻ Assumption: global quality sets the stage for instructional quality to be possible

□ CLASS and ECERS-R administered at all levels

- ◻ Assumption: both measure distinct aspects of quality

More to learn!

Summary

- Making sure that individual component measures function well is paramount to the functioning of the overall rating
    - Need to consider implementation consistency, how measures are collected, and reliability of measures

- Constructing a valid QRIS is hard!
    - Conducting developmental evaluations with child outcomes can help states examine components and adjust scoring approaches
    - Collecting base/raw level data is important for this

Summary

- Cut-scores are a critical part of a QRIS
  - Careful consideration should be made regarding cut-scores as slight adjustments can effect ratings, funding, and the ability to detect relationships with child outcomes

- Coming attractions: We still have more to learn about how best to measure aspects of quality such as:
  - Curriculum implementation
  - Assessment use
  - Family partnerships

Questions?

E-mail: diana.schaack@ucdenver.edu

Thank you for joining us!

CA-QRIS Members may disconnect now.

Rating Matrix Workgroup members, please stay on the phone for a planning call.