Let’s Talk

Host - Debi Mathias
Director, QRIS National Learning Network

dmathias@buildinitiative.org
www.buildinitiative.org
www.qrisnetwork.org
QRIS Ratings and Outcomes: Psychometric Issues and Validation

Sandy Soliday Hong
Frank Porter Graham
University of North Carolina, Chapel Hill
sandy.hong@unc.edu

Terri J. Sabol
School of Education and Social Policy
Northwestern University
terri.sabol@northwestern.edu
Overview

- Overview of QRIS Validation
- Validating Existing QRIS: Simulation Study
- Drawing Upon Psychometric Theory to Inform QRIS Ratings
- BUILDing Next Steps Together
Overview of QRIS Validation
Validation of QRIS

Validating QRIS is an important step in development, implementation, and evaluation stage in QRIS.

Definition of validation for QRIS:

*Is the QRIS performing as it was intended?*
Race to the Top- Early Learning Challenge and Validation of QRIS

- $1 billion in 3 rounds; 20 states

Goals
- State Systems
- Accountable Programs
- Child Outcomes
- Evaluation/validation

Map showing states: Round 1: CA, CO, GA, MN, NC, OR, WA; Round 2: CA, CO, GA, MN, NC, OR, WA; Round 3: CA, CO, GA, MN, NC, OR, WA.
19 QRIS conducting a validation study

9 QRIS conducting validation studies that include children’s outcomes

19 QRIS conducting an implementation study
Validating QRIS: Challenges and Opportunities

- QRIS are a multiple dimensional system
- Systems present challenges for validation
Validating QRIS: Challenges and Opportunities

Goal or Mission of QRIS
Population
Ratings
Market Forces
Improvement Supports
Child Outcomes
Numerous, important goals of QRIS

Growing interest and investment in validating ratings with child outcomes

Use child outcomes as an example of whether QRIS are related to intended outcomes
Multiple aspects of quality

Key question:

What is a high quality program/classroom?
Important theoretical and empirical work on the underlying model of early childhood quality

- Indirect Quality Indicators
  - Teacher education
  - Director education
  - Ratio/Class size
  - Curriculum

- Direct Quality Indicators
  - Age appropriate activities
  - Intentional learning activities
  - High quality teacher-child interactions

- Child Outcomes
Almost all QRIS include **direct quality indicators**
- Learning environment, teacher-child interactions

Almost all QRIS include **indirect quality indicators**
- Staff qualifications, curriculum, staff-child ratio

QRIS Ratings in 2015

BUILD Initiative & Child Trends, 2014
QRIS Ratings in 2015

- Based on evidence from quality research

- Quality indicators are combined into a rating
  - Point system, building block, or hybrid

- Decisions on how combined scores impact ability of QRIS rating to accurately describe quality

- Little evidence on validity of ratings
Validating Existing QRIS: Simulated QRIS Study

Simulated QRIS Study

- Simulate current state rating systems with secondary dataset
- Explore how tightly aligned QRIS quality indicators and rating are with intended outcomes
- Examine relation among quality indicators and ratings with child outcomes
Simulated QRIS Study Dataset

- The National Center for Early Development and Learning (NCEDL) Multi-State Study
  - 2001-2002
  - Six states
  - Funded by the U.S. Department of Education

- State-Wide Early Education Programs Study (SWEEP)
  - 2003-2004
  - Five states
  - Funded by private foundations
Simulated QRIS Study
Selection of Individual Quality Indicators

- Indicator 1: Staff Qualifications
- Indicator 2: Ratio and Group Size
- Indicator 3: Family Partnership
- Indicator 4: Environment (ECERS)
- Indicator 5: Teacher-child interactions (CLASS)
Results
Individual quality indicators and child outcomes

* \( P < .05 \)
Results
Individual quality indicator star ratings and child outcomes

- High-star versus low-star contrast

* P<.05
Simulated QRIS Study
Creating Ratings

Direct Quality Indicators
- Age appropriate activities
- Intentional learning activities
- High quality teacher-child interactions

Indirect Quality Indicators
- Teacher education
- Director education
- Ratio/Class size
- Curriculum

Rating

Child Outcomes
## Simulated QRIS Study
### State Rating Structure

<table>
<thead>
<tr>
<th>State/Region</th>
<th>Point</th>
<th>Building Block</th>
<th>Hybrid</th>
<th>Number of Stars</th>
<th>% in highest level</th>
</tr>
</thead>
<tbody>
<tr>
<td>State A</td>
<td>X</td>
<td></td>
<td></td>
<td>5</td>
<td>16</td>
</tr>
<tr>
<td>State B</td>
<td>X</td>
<td></td>
<td></td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>State C</td>
<td>X</td>
<td></td>
<td></td>
<td>5</td>
<td>85</td>
</tr>
<tr>
<td>State D</td>
<td></td>
<td>X</td>
<td></td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>State E</td>
<td></td>
<td></td>
<td>X</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>State F</td>
<td></td>
<td></td>
<td>X</td>
<td>4</td>
<td>78</td>
</tr>
<tr>
<td>State G</td>
<td></td>
<td></td>
<td>X</td>
<td>5</td>
<td>25</td>
</tr>
<tr>
<td>State H</td>
<td></td>
<td></td>
<td>X</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>State I</td>
<td></td>
<td></td>
<td>X</td>
<td>5</td>
<td>84</td>
</tr>
</tbody>
</table>

All include four quality indicators:

1. Staff Qualifications
2. Environment
3. Family Partnership
4. Ratio and Group Size
Results
Simulated state star ratings and child outcomes

High-star versus low-star contrast

* P<.05
Implications

- Need a clearly articulated theory of change and align to rating system
- Need more front end work about how to structure ratings
Let’s Talk
Drawing Upon Psychometric Theory to Inform QRIS Ratings: Meta-Analysis Study
Purpose: Examine QRIS from a psychometric perspective

- Explore how research and evaluation can assist in creating and refining quality rating systems,
- Discuss some of the challenges faced in attempting to implement such processes

Psychometric issues related to scale development can be useful
New meta-analysis examined a simplified point approach based on professional standards, and asked

- Do selected QRIS structural indicators predict process quality?
- Do selected QRIS structural and process indicators predict child outcomes?
**QRIS: Psychometric Issues**

- **Item Selection**: process for determining which quality measures to include in a rating system

- **Dimensionality**: whether rating scales are uni-dimensional

- **Item Scoring**: process for determining how quality variables are categorized into indicators

- **Reliability Standards**: linking standards to how data are used
Item Selection: process for determining which quality measures to include in a rating system

- Ideally: items are measuring different, but related, components and are predictive of the underlying construct
- QRIS: items selected based on “evidence” they measure quality
  - Level of evidence varies widely
    - Typically focus on whether any study produced “significant” associations
    - Ideally focus on magnitude of association in the most rigorous studies
**Item Selection**: All items should correlate positively with general quality constructs

- *Examples with strongest research support*
  - Curricula
    - Relatively consistent evidence - presence of curricula
    - Growing evidence – focused, sequenced, intensive curricula
  - Ratio/group size
    - Well established professional guidelines
    - Many studies – most supporting professional guidelines
  - Teaching staff education
    - Well established professional guidelines
    - Many studies – over half supporting professional guidelines
Item Selection: All items should correlate positively with general quality constructs

- Examples – with consensus on importance, but less evidence for specific guidelines
  - Director education and experience
    - Consensus regarding importance and clear professional guidelines
    - Fewer studies
  - Family involvement
    - Consensus regarding importance, but no clear professional guidelines
    - Fewer studies – very mixed in focus and results
  - Diversity and dual language learners
    - Consensus regarding importance, recent professional guidelines
    - Fewer studies – very mixed in focus and results
**Dimensionality**: whether rating scales are uni-dimensional

- **Ideally**: all items are positively correlated, suggesting increases on any item represents the same thing in the scale score.

- **QRIS**: items are selected because they represent different, often uncorrelated, aspects of quality.
  - Low to moderate correlations among quality indicators.
  - Low to moderate correlations between quality indicators and either classroom quality measures or child outcomes.
QRIS: Psychometric Issues

- **Item Scoring**: process for determining how quality variables are categorized into indicators
  - **Ideally**: all items are optimally scored (scores are as predictive of the underlying construct as possible)
  - **QRIS**: cut-offs are chosen based on
    - Professional guidelines or evidence from research studies
    - Ensuring programs can progress through ratings to make enrolling in QRIS more attractive
**QRIS: Psychometric Issues**

- **Reliability Standards**: linking standards to how data are used
  - Ideally: reliability focuses on how the instrument is used and ensuring it is as accurate as possible as used
  - Should consider multiple sources of variability
    - Rater
    - Time
    - Classroom within centers

- **QRIS**: professional standards of reliability (often research standards) are applied
  - CLASS or ECERS: 85% agreement within one but need to know accuracy of placing programs into QRIS levels rather than accuracy of total score in general
Meta-Analysis Study Datasets

- Data from 6 studies
  - Head Start
    - FACES 2006
    - FACES 2009
  - Pre-kindergarten evaluations
    - Georgia Pre-k
    - North Carolina Pre-K
  - Mixed
    - NCRECE evaluation of MyTeachingPartner
    - ECLS-B
Meta-Analysis Study Results: Psychometric Properties

- **Reliability:**
  - Data collected with research standards
  - More work to be done on reliability of data collection with observational measures

- **Dimensionality:**
  - Factor analysis: 2-3 factor in all 6 studies
    1. Teacher and Direction education, & curriculum
    2. Ratio/group size
    3. Observed process quality (CLASS, ECERS)
  - Internal consistency: Modest to moderate
  - Summary – even classroom quality indicators appear to be multi-dimensional
Item Selection:
- Used indicators with best evidence
- They related to both process quality and child outcomes (modest to moderate associations)

Item Scoring
- Used well established guidelines to create cut-points
- Turning the continuous indictors into ratings did not markedly change associations with process quality and child outcomes
- Suggests these guidelines appear to retain the relevant information in the indicators
Meta-Analysis Study Results: Indirect Predictors of Child Outcomes

- T Ed
- Ratio
- Curric
- D Ed
- Group Size

Legend:
- Language
- Pre-Literacy
- Math
- Social Skills

* Indicates significance.
Meta-Analysis Study Results:
Overall QRIS Ratings as Predictors of Child Outcomes

* p<.05
Conclusions

- Limitations
  - Validity: Secondary data analyses of research data
  - Generalizability: Most studies were of Head Start or Pre-Kindergartens

- Tentative conclusions
  - When careful attention is paid to the psychometric properties of the rating scale, the QRIS approach can describe the information about program quality in a manner that is related to gains in child outcomes
BUILDing Next Steps Together
Overall Summary

- **Translational Research**
  - Established research base → inform practice & policy

- Individual indicators relate to children’s learning

- Once indicators are aggregated together, associations with child outcomes are diminished
Questions for Larger Group

- What research would be most useful for your state?

- How can research best inform your state’s rating structure?
Debi Mathias, Director QRIS NLN
dmathias@buildinitiative.org
BUILD Initiative
www.buildinitiative.org
QRIS National Learning Network
http://qrisnetwork.org/