

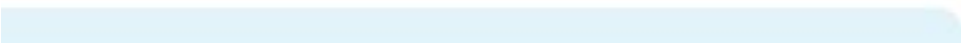
Validating Quality Rating and Improvement Systems

Webinar
March 15, 2012



INQUIRE

Quality Initiatives Research and Evaluation Consortium



Validating Quality Rating and Improvement Systems

Webinar
March 15, 2012



Welcome

- Ivelisse Martinez-Beck
 - Child Care Research Coordinator, Office of Planning Research and Evaluation

Welcome

- Shannon Rudisill,
 - Director, Office of Child Care



INQUIRE

Quality Initiatives Research and Evaluation Consortium

Welcome

- Kathryn Tout
 - Co-Director of Early Childhood Research, Child Trends

Goals of the Webinar

- Introduce a framework for QRIS validation
- Describe real examples of state validation efforts
- Highlight challenges and lessons learned
- Offer guidance on developing an individualized state plan for QRIS validation

QRIS Validation Panelists

- Gail Zellman, RAND Corporation
- Kelly Maxwell, University of North Carolina
- Michel Lahti, University of Southern Maine
- Jim Elicker, Purdue University
- Kim Boller, Mathematica Policy Research
- Kathryn Tout, Child Trends

What is QRIS Validation?

- An ongoing, iterative process that assesses whether design decisions about program quality standards and measurement strategies are producing meaningful and accurate ratings
- QRIS validation studies assess whether rating components and summary ratings can be relied on as accurate indicators of program quality
- Validation studies can identify needed changes and support continuous quality improvement

Why is Validation Important?

- Promotes increased credibility and support for the QRIS
 - Parents can rely on ratings in selecting care
 - Providers more willing to participate
- Supports effective deployment of limited rating resources (measuring only those things that contribute to quality)
- Promotes efficient use of limited QI resources
 - Technical assistance can target key aspects of care
 - Providers can use ratings to target QI efforts

What is QRIS Validation?

- A complex iterative process
- Relies on multiple sources of evidence
 - Expert judgments of degree to which measures capture key quality components
 - Scores on different measures of the same concept
 - Patterns of relationships
 - Across scores on different measures
 - Among the items within a measure

What is QRIS Validation?

- Four approaches may be used
 1. Examine validity of key underlying concepts
 2. Examine the psychometric properties of measures used to assess quality
 3. Assess the outputs of the rating process
 4. Relate ratings to expected child outcomes
- Approaches vary in terms of timing, cost, difficulty
- Approaches are not rigid; may overlap in time and goals

1. Examine the Validity of Key Underlying Concepts in QRIS

- Assesses whether basic concepts included in QRIS rating are the “right” ones by examining level of empirical and expert support
- Addresses questions like:
 - Do the rating components capture the key elements of quality?
 - Is there sufficient empirical support for including each component?
- Ideally conducted prior to QRIS implementation

1. Examine the Validity of Key Underlying Concepts in QRIS

- Data needed
 - Empirical literature on relationship of components to high quality care
 - Expert views
- Analysis methods
 - Synthesis of available data to determine level of support for each component
 - Consensus process

Example: Indiana

- Indiana Paths to QUALITY
 - Purdue University
- Questions: What does research tell us about
 - Whether the QRIS components and levels result in increasing quality of programs?
 - Whether the QRIS will improve developmental outcomes for children?

Example: Indiana

- Methods:
 - Comprehensive review
 - Classified each QRIS indicator as having “some,” “moderate,” or “substantial” evidence
- Found “substantial” evidence for 75% of indicators

Example: Georgia

- Georgia Department of Early Care and Learning
- Question: What are the key indicators of quality?
- Methods for Addressing:
 - Stakeholder group
 - Expert review
 - Crosswalk with other program standards
 - Statewide study of quality of care
- Validation helped identify set of indicators included in pilot QRIS

Example: Kentucky

- Kentucky STARS for KIDS NOW
 - Child Trends
- Question: How do current standards align with existing quality frameworks?
- Methods:
 - Crosswalk comparison of standards & frameworks
- Confirmed some standards and identified possible gaps

2. Examine the Psychometric Properties of the Measures Used to Assess Quality

- Assesses whether component measures and overall ratings perform as claimed and expected by theory
- Addresses questions like:
 - Do component measures which claim four scales actually have four scales?
 - Do measures of similar concepts relate more closely to each other than to other measures?
 - Do different cut scores produce better distributions or more meaningful distinctions among programs?

2. Examine the Psychometric Properties of the Measures Used to Assess Quality

- Data needed
 - Rating data from participating programs
 - Data on additional quality measures
- Analysis methods
 - Factor analyses of some measures
 - Correlations among components
 - Correlations of selected components with other measures of quality

Example: Maine

- Maine – Quality for ME
 - University of southern Maine
- Random selection of programs by type and Step Level over time. (3 year period)
- Data sources for the validation study
 - On site observations using the ERSs.
 - Confidential staff questionnaire
 - Anonymous parent questionnaire
 - Administrative Data:
 - QRIS Enrollment Data (self-report thru web based system)
 - ME DHHS Licensing Data
 - ECE Training and Technical Assistance Data (Registry Data)

Example: Maine

- Supports to parents is measured by the QRIS (provider self report).
- Do parents' survey reports of services align with providers' self report?
 - “Did you receive this support/service?”
 - Higher step levels more likely to provide more supports.
 - Differences found across family child care, Head Start centers
- Use of Emlen (2000) 15-item scale – perceptions of quality:
 - Good reliability
 - Did not discern differences in total mean scores by step level.
 - Step level rating did correlate with 6 individual items.
 - Healthy place for my child / Caregiver knows a lot about children / Lots of creative activities going on / Interesting place for my child / Child is treated with respect / Caregiver is open to new information
 - Found total mean score differences by program types.

Example: Maine

- Staff development and supports and resources for staff are measured in the QRIS (provider self report)
- Do staff/provider surveys of job stress, demands and resources align with step level?
 - Higher step level program standards focus more on staff development, supports and resources for staff.
 - Secondary data available from Registry (Education / Training)
- Measures:
 - Job Stress Inventory (Curbow et al, 2000): Job Demands, Job Control and Job Resources subscales – no relationship to step levels.
 - Differences were found by type of program.

Example: Kentucky

- Question: What strategy for combining indicators will produce valid rating levels?
- Method: Use data from the existing STARS rating and additional data from a survey of providers to simulate alternative ratings
 - Current structure is a block design
 - Alternative models included a point system and a hybrid model using blocks and points
- “New” and existing ratings were compared to understand the impact of different structures

3. Assess the Outputs of the Rating Process

- Examines program-level ratings scores to assess rating distribution and relationship of ratings to other quality measures
- Addresses questions like:
 - Are providers that received 4 stars actually providing higher quality care than those that earned 3 stars?
 - Do rating distributions for programs of different types, e.g., center vs. home-based vary?
 - Are cut scores and combining rules producing appropriate distributions?

3. Assess the Outputs of the Rating Process

- Data needed
 - Program-level ratings from participating programs
 - Data from additional quality measures
- Analysis methods
 - Examination of rating distributions by program type
 - Correlations of program ratings with other measures
 - Changes in rating distributions over time

Example: Indiana

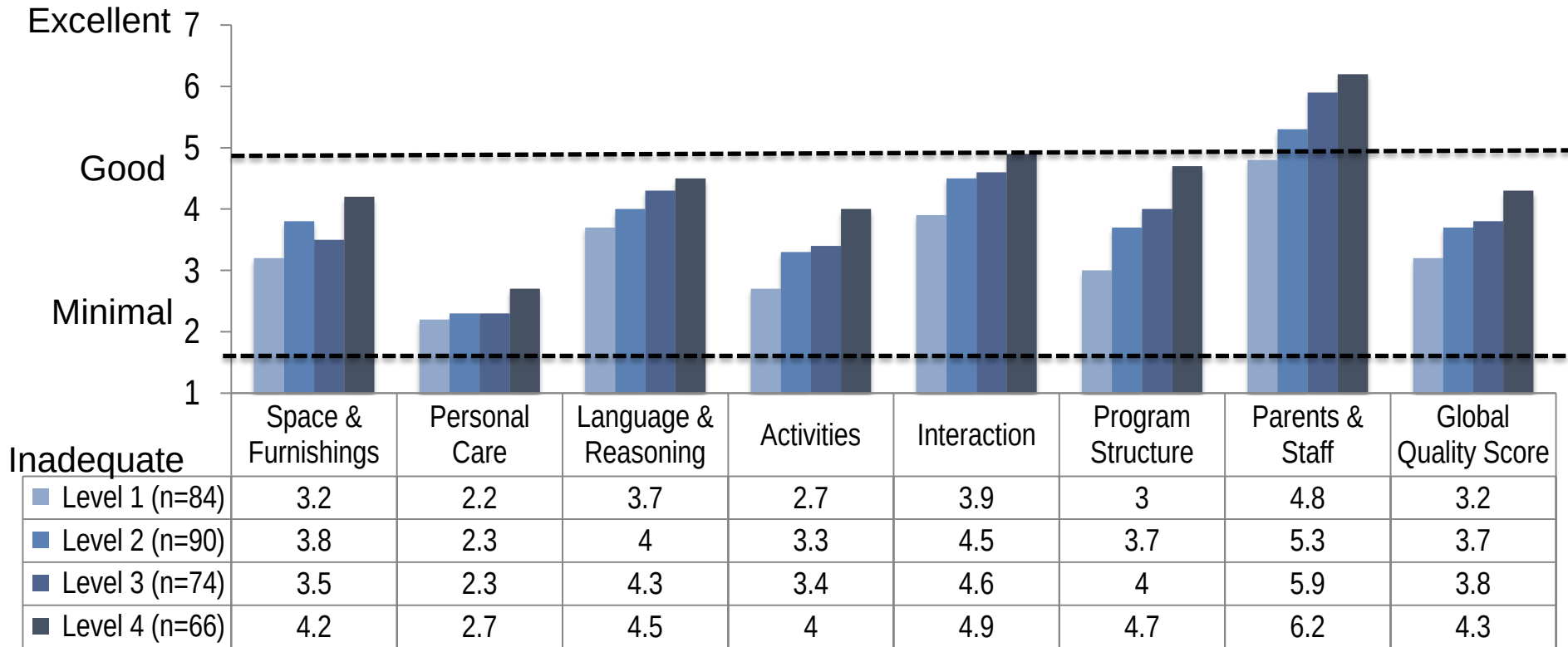
- Question:
 - Do Indiana PTQ level ratings correlate with quality, as measured by established measures: ECERS-R, ITTERS-R, FCCERS-R?

Example: Indiana

- Method:
 - ERS quality assessments using the scales in a stratified random sample of Level 1, 2, 3, and 4 PTQ-rated child care providers

Key Findings:

All Providers: Average ERS scores by PTQ level (n=314)



(Correlation between ERS global scores and PTQ ratings: $r = .45^{***}$)

Example: Indiana

- How findings are being used:
 - Findings were presented to stakeholders in 11 regional meetings around the state
 - Series of research briefs for general public and policy makers in development
 - Follow-up study: Identify lowest-rated quality areas in the evaluation sample

Example: Indiana

- Issues:
- What should be the quality validation standard(s) for each state?
 - Based on national research-validated measures of quality? (Accreditation; research measures)
 - Or based on local definitions of quality?
- What are the criteria for acceptable quality validation evidence?

Example: Indiana

- Question:
 - Among Level 3 and 4 PTQ rated providers, what are the quality areas most in need of improvement?

Example: Indiana

- Detailed examination of ERS data for Level 3 and Level 4 rated providers in the evaluation sample
- Separate analyses for ECERS-R, ITERS-R, and FCCERS-R
- Identified ERS items with means < 4 , and ERS items with means < 3

Example: Indiana

- **Level 4 Providers had 10 ECERS-R items that averaged below 4.0:**
(*7 items that averaged below 3)
- **Level 3 Providers had 15 ECERS-R items that averaged below 4:**
(* 7 items that averaged below 3)

Example: Indiana

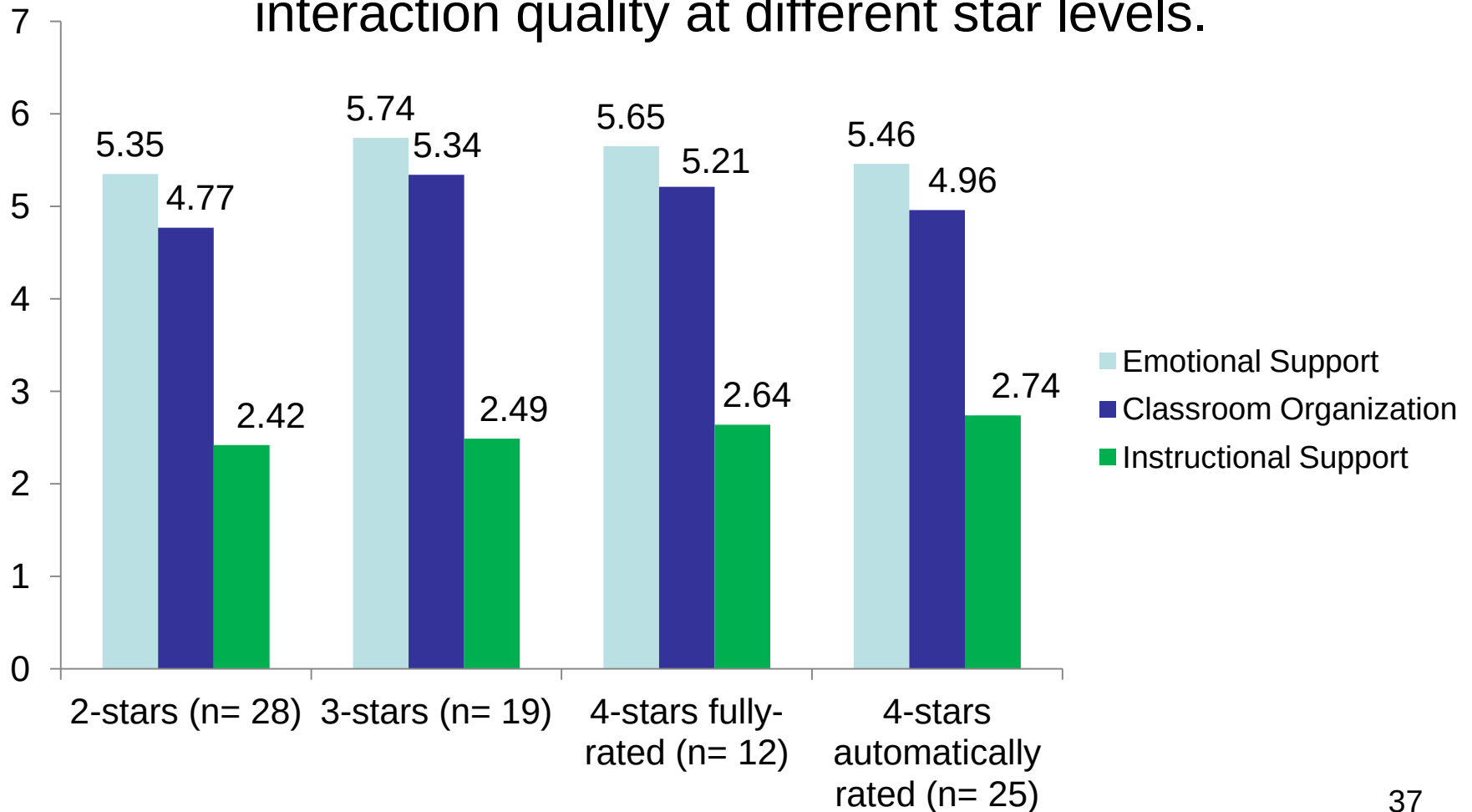
- Findings presented to PTQ committees: evaluation, standards revision, provider resources
- Data are used as a guide. ERS data alone will not drive revisions.
- Committees will consider changes in PTQ standards, rating procedures, and T/TA based on these findings.

Example: Minnesota

- Parent Aware
 - Child Trends
- Question: Does the quality of teacher-child interactions differ by star level?
- Method: Examine patterns of scores on the CLASS (which is included in the rating tool - points are awarded for higher scores)

Example: Minnesota

There were no differences in observed teacher/child interaction quality at different star levels.



No significant differences at initial ratings.

Example: Minnesota

- Challenges:
 - Small numbers of programs overall and limited numbers of programs at the lower star levels
 - CLASS was conducted only in center-based preschool rooms. What about other age groups?
- Findings were used in the revision of Parent Aware standards and statewide rollout

4. Relate Ratings to Expected Child Outcomes

- Examines the extent to which exposure to higher quality providers is associated with better child functioning
- Addresses questions like:
 - Do higher-rated programs produce better learning outcomes?

4. Relate Ratings to Expected Child Outcomes

- Data needed
 - Rating data from participating programs
 - Assessments of child functioning
- Analysis methods
 - Examine statistical relationship between ratings and child outcomes
 - Rigorous analytic methods should be used to account for selection factors and sampling bias

Example: Minnesota

- Question: Do gains in children's school readiness vary by star level or quality component?
- Method: 700 four-year olds were recruited from 138 QRIS-rated programs (up to 6 from selected classrooms and 4 from family child care programs). Low-income children were prioritized.
 - In the fall and spring of the year before kindergarten, children completed direct assessments of expressive and receptive vocabulary, phonological awareness, print knowledge, and early math skills
 - Teachers/caregivers completed assessments of children's social-emotional development and approach to learning

Example: Minnesota

- Fall to spring gains on measures were calculated and compared across star rating levels (combining children from 1- and 2-stars) and quality categories using multilevel models
 - Four quality categories: Family Partnerships, Teacher Training and Education, Tracking Learning, and Teaching Materials and Strategies
- No consistent evidence that children's gains varied by star rating or by points earned in the different quality categories

Example: Virginia

- Virginia Star Quality Initiative
 - University of Virginia
- Questions:
 - Do ratings of pre-kindergarten programs relate to children's early literacy skills as they enter kindergarten?
 - Do gains in early literacy from pre-kindergarten to kindergarten relate to program ratings?

Example: Virginia

- Method: Multi-level models were estimated using rigorous controls at the child- and center-level and community fixed effects
- No significant differences in literacy skills at kindergarten entry by star level
- Children in 3- and 4-star programs did show more significant growth in literacy skills in the year before kindergarten compared to children in 2-star programs
- Differences in growth were not sustained into kindergarten

Issues with Inclusion of Child Outcomes in QRIS Validation Studies

- Studies are costly when they include direct assessments
- Limited measures to use with children who are English Language Learners
- Difficult to measure attendance and exposure to a program
- Studies are not measuring how the QRIS affects children; they measure associations between ratings and measures of children's development
- Methods should account for nesting of children within programs and control for selection factors
- See new paper by Zellman and Karoly

Approaching Validation with a Plan

- Given complexity, useful to develop a plan early in the process, before QRIS implementation
 - Thinking about validation may help in the design phase
 - Some validation data can be collected as part of ratings or other QRIS activities
- Plan ideally should include all four approaches

Validation Plan Considerations

Approach	Timing	Cost issues	Getting by
Examine the Validity of Key Underlying Concepts	Ideally, do before implementation Should take just a few months	Relatively inexpensive	Can rely on other states' efforts for many measures
Examine the Psychometric Properties of the Measures Used to Assess Quality	Must wait until ratings occur Can conduct several studies using same data set	Depends on data quality and amount of analysis Additional measures will increase costs	Can rely to some extent on available resources
Assess the Outputs of the Rating Process	Must wait until ratings occur Can conduct several studies using same data set	Depends on data quality and amount of analysis Additional measures will increase costs	This work is system-dependent but lessons learned about structure and cut-points can be shared
Relate Ratings to Expected Child Outcomes	Best to delay until ratings process stable and sufficient programs have been rated	Child data collection costs very high Some agencies may collect these data	Requires significant funds and expertise; sampling children and programs reduce costs

Five Takeaways

1. QRIS validation is a process
2. A validation plan charts a course for short-term and long-term validation activities
3. Usually the four validation approaches are assessed sequentially, but they can overlap in time
4. Validation is a measure of a successful QRIS and requires assessment at all stages of implementation
5. Validation tools and technical assistance resources are available

QRIS Validation is a Process

- Validation is not a destination
- Validation activities can be done on their own or as part of a research and evaluation agenda
- Validation assesses whether the system and its measures of quality are working as designed
- Never too late to prepare and implement a validation plan

Benefits of Planning for the Long Haul

- A shared, comprehensive, phased plan for validation can help weather funding uncertainties
 - Stepping outside of immediate budget constraints encourages deeper thinking and setting of both short-term and long-term goals
 - Getting cross agency and stakeholder buy-in leverages support for a comprehensive long-term plan
- Long-term validation, implementation, and outcome evaluation planning reveals synergies that may reduce future data collection and analysis costs

The Four Validation Approaches Build on Each Other

- Assessment of the validity of key concepts is foundational for the other approaches
- Assessment of psychometric properties is often done as part of an output or outcome assessment
- Ideally, assessment of outputs comes before assessment of child outcomes

Validity Assessment is Important at Each Stage of QRIS Implementation

- Pilot and scale-up validation usually focuses on the first two approaches
- The early operation stage (2-5 years) may focus on any of the four approaches, but the first three approaches predominate
- Mature stage (>5 years) validation assessments provide the opportunity to refresh standards and measures and inform system changes
- When significant changes are made to the ratings or measures of quality, that should trigger a validity assessment

Validation Tools and Resources

- QRIS Evaluation Toolkit (Lugo-Gil et al. 2011)
 - Includes a section on validation
 - Provides examples of study designs to answer validation questions
 - Provides examples of the measures used in state studies
 - Includes costs for actual studies as well as features of a strong request for proposals
- Recently released state reports
- Forthcoming OPRE validation brief
- TA centers
- INQUIRE members

Validation Study: QRIS Evaluation Toolkit

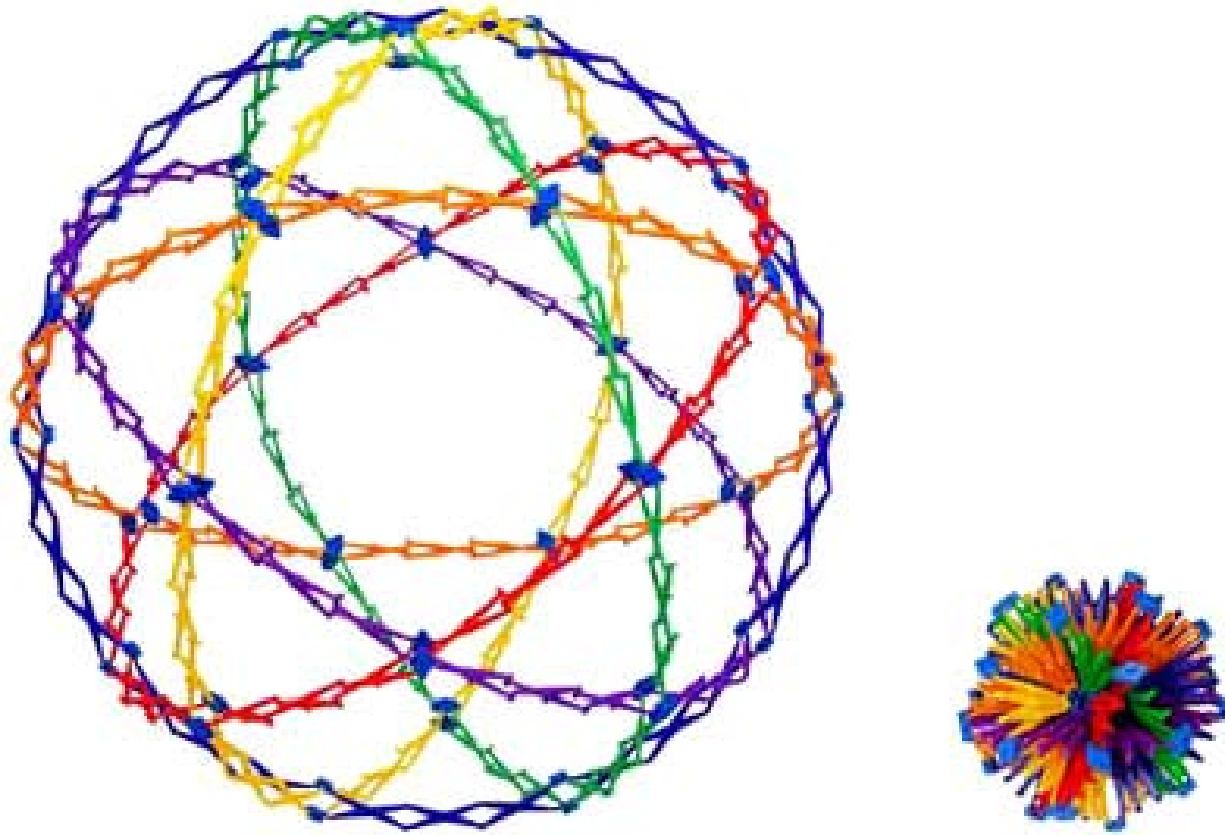
A validation study assesses the degree to which the quality standards component of the QRIS reflects meaningful quality levels that are linked to desired outcomes.

What is this approach?	It is a way to ask critical questions about the tools used in a QRIS and how they are functioning
What can be learned with this approach?	Whether the quality ratings actually mean something important to programs, parents, and children
What research questions can be answered with this approach?	• Do the quality standards reflect the current research base? • Do the quality standards represent distinct areas that do not overlap with other standards in the QRIS? • Can improvement be detected in the quality levels? • Are the quality levels related to children's functioning?
What are the key factors to consider when using this approach?	• The timing of the validation study and how findings can inform QRIS design and implementation • The degree to which the QRIS includes a representative sample of providers from the communities served by the QRIS

Ongoing Validation Assessment Issues

- Characteristics and density of providers enrolled in the system
 - In voluntary systems that have over-representation of higher quality settings or a small number of participants, results may be skewed
- Validation does not take the place of implementation and outcome evaluation
- Cost and fluctuation in commitment to validation efforts

Validation Efforts Can Be Scoped to Available Resources



Questions and Discussion

Recording and Resources will be available on:

- <http://researchconnections.org>

**Child Care & Early Education
RESEARCH CONNECTIONS**

Presenter Contact Information

- Ivelisse Martinez-Beck – ivelisse.martinezbeck@acf.hhs.gov
- Shannon Rudisill – shannon.rudisill@acf.hhs.gov
- Kathryn Tout – ktout@childtrends.org
- Gail Zellman – zellman@rand.org
- Kelly Maxwell – maxwell@unc.edu
- Michel Lahti - mlahti@usm.maine.edu
- Jim Elicker - elickerj@purdue.edu
- Kimberly Boller – kboller@mathematica-mpr.com

Reports

- Indiana - Evaluation of Paths to Quality
- <http://www.cfs.purdue.edu/cff/publications/publications.html>
- Minnesota – Evaluation of Parent Aware
- <http://www.melf.us/>
- Kentucky
- <http://www.kentuckypartnership.org/starsevaluation.aspx>
- Maine
- Contact Michel Lahti for further information
- QRIS Evaluation Toolkit
- http://www.acf.hhs.gov/programs/opre/cc/childcare_quality/gris_toolkit/gris_toolkit.pdf
- INQUIRE products
- http://www.acf.hhs.gov/programs/opre/cc/childcare_technical/index.html

Reports

- Virginia
- Contact Terri Sabol for further information
- terri.sabol@northwestern.edu
- RAND (Zellman and Karoly) report on child assessments in QRIS
- http://www.rand.org/pubs/occasional_papers/OP364.html
- http://www.rand.org/pubs/research_briefs/RB9639.html