

Validation of Quality Rating and Improvement Systems for Early Care and Education and School-age Care

**Gail Zellman
Rick Fiene**

November 17, 2011

What is Validation?

- An ongoing process that assesses the degree to which evidence and theory support conclusions from assessments conducted in a specified context
- For QRISs, validation studies assess whether rating components and summary ratings can be relied on as accurate indicators of program quality
- Validation also addresses whether key system components are operating as expected

Concept of Validation Has Changed

- In the past, five kinds were delineated
 - Face validity
 - Construct validity
 - Content validity
 - Predictive validity
 - Concurrent validity
- No longer five types; validity is now considered a single concept

However, goal remains the same: to build a case that measures accurately reflect the concepts they purport to measure

How Does Validation Differ from Evaluation?

- Validation focuses on measurement tools
 - Seeks to determine if those tools measure what they purport to measure
 - Goal is to improve measures and other system features
- Evaluation focuses on whether a specific program or intervention is effective in reaching its objectives
- Some QRIS evaluations may also address validity questions
 - E.g., finding that participating programs improve their quality suggests QRIS is working
 - Finding associations in the same data between rating level and amount of improvement supports underlying QRIS logic model and rating system

A valid rating system does not ensure that an evaluation will find effects. But an invalid rating system ensures that any evaluation results will not be meaningful.

Why Does Validation Matter in QRISs?

- Assessment and rating are key aspects of QRISs; these systems rely on ratings
- Accurate assessments are essential to deliver on the promises of QRISs
 - Parents can rely on ratings in selecting care
 - Providers can use ratings to target QI efforts
 - Technical assistance can target key aspects of care

Why Should Policy Makers Care about Validation?

- Increased support for the QRIS (if better-rated providers are better)
- Effective deployment of limited rating resources (measuring only those things that contribute to quality)
- Appropriate use of limited QI resources (if quality is well-measured and components matter)
- Avoidance of legal challenges
 - Assessments have increasingly high stakes attached to them; providers care about their rating and may question it
 - Validation studies provide support for ratings

What Does it Mean to Validate a QRIS?

- A complex iterative process
- Relies on multiple sources of evidence, e.g.,
 - 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 Does it Mean to Validate a QRIS?

- Four approaches may be used
 - Examine validity of key underlying concepts
 - Examine the psychometric properties of measures used to assess quality
 - Assess the outputs of the rating process
 - Relate ratings to expected child outcomes
- Approaches vary in terms of timing, cost, difficulty

Examine the Validity of Key Underlying Concepts

- 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
- Example: Indiana’s *Paths to Quality* validation

Examine the Validity of Key Underlying Concepts

- 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 components
 - Consensus process

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?
- Examples: Gordon et al., Perlman et al. on the *ECERS-R*

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

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?
- Example: Maine's *Quality for ME* validation; Karoly and Zellman's virtual pilot work

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

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?
- Example: Colorado's *Qualistar Early Learning QRIS* (Zellman et al); Missouri (Thornburg et al)

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
 - With random assignment, can estimate causal effect of QRIS on child outcomes

Approaching Validation with a Plan

- Given complexity, useful to develop a plan early in the process, before 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 system-dependent but data probably available for an intern or grad student
Relate Ratings to Expected Child Outcomes	Best to delay until full implementation and time for change	Child data collection costs very high. Some agencies may collect these data	Requires significant funds and expertise