

Issues and Challenges: Measuring Quality in Quality Rating Systems

Description

The goal of this session was to understand the complexity of issues related to state-level quality assessments. Participants discussed implications of measuring quality in quality rating systems across the range of settings and age groups (infant/toddlers, preschool, school age); use of research; and challenges of the currently available measures of quality and their applicability to quality rating systems (QRS).

Facilitator

David Edie, National Child Care Information Center

Discussants

- Kathy Thornburg, Midwest Child Care Research Consortium
- Mary Roach, Wisconsin Child Care Research Partnership
- Jamie Gottesman, Ohio Department of Job and Family Services

Scribes

- Linda Mayo Willis, University of Nebraska
- Sheri Azer, National Child Care Information Center

Issues

- How do states operationalize their definitions of quality?
- What measures are states using to monitor quality?
- What issues and challenges do states face in measuring quality?
- How has or could research support states' efforts?
- What policy implications are emerging related to quality rating systems (particularly monitoring and accountability)?
- What are the implications of measurement of quality on provider and/or child outcomes?

Documents in the Session Folder:

- Quality Rating Systems: Definitions and Statewide Systems
- Common Categories of Criteria used in State Quality Rating Systems
- Quality Rating Systems and the Impact on Quality in Early Care and Education

Matthew Richardson 10/30/06 10:30 AM
Deleted: 10

Discussion Notes

Summary of issues discussed by panel members and participants:

- To what extent should detailed, on-site assessments of quality determine the tiers or star levels in a quality rating system? Some states use environment ratings scales to determine star levels and some use them or other research tools to verify ratings drawn from administrative data or less costly assessments.
- Participants discussed accreditation by NAEYC (National Association for the Education of Young Children) as a research-based standard for quality measurement as well as differences in the effectiveness of the old and new systems.
- Participants discussed whether a QRS includes different types of providers – such as centers, family child care (FCC), and school-age care (SAC) – and whether equivalent measures are used across program types. How to include family child care (FCC) homes, which often have different qualifications and learning environments, was also discussed. States represented in the session agreed that they created their systems to be inclusive.
- Participants discussed how states verify that their quality system is valid and how states measure child outcomes.
- There was discussion on whether to base QRS ratings on ECERS (Early Childhood Environmental Rating Scale) scores or whether to use ECERS to validate the rating system itself.
- Discussion questions included: How do you know if QRS is a worthy investment? How do you prove that it is? Is money better invested in direct quality improvement?

Notes from the General Discussion

Overview on Quality Rating Systems

- 13 states are running statewide QRS that include child care and Head Start.
- 25 states are looking at developing QRS.
- This is a new policy event in early care and education.
- Policy development is moving faster than research.

Matthew Richardson 10/30/06 10:30 AM
Deleted: 10

- The *Research Connections* website has good information on policy-related research.
- Common areas in QRS standards:
 - * Program Development
 - * Learning Environment
 - * Licensing status
 - * Parent involvement
 - * Staff competence
 - * Program evaluation
- Environmental Ratings and Tiered Reimbursement in QRS:
 - * 9 states are using ECERS-R (ECERS-Revised).
 - * There is considerable variation in how ECERS is used, frequency of assessment, percent of classrooms, and how scores are used.
 - * 30 states have tiered reimbursement policies based, at least in part, on environmental ratings.
 - * Some states limit environment ratings to top tiers.
- Oklahoma, North Carolina, and Tennessee are evaluating their quality rating systems.

Mary Roach

The Wisconsin QRS criteria were developed with a research focus and are not yet implemented:

- Simple:
 - * To implement
 - * To operate
- Valid and realistic:
 - * Reliable measure
 - * Measures quality
 - * Realistic for programs to progress from one level to the next
- Efficient:
 - * Reasonable to administer
 - * Reasonable cost

Matthew Richardson 10/30/06 10:30 AM
Deleted: 10

- Research to policy:
 - * Tough to merge
 - * Difficult to translate research
 - * QRS model tries to simplify complex research information
- Continuum of quality:
 - * Proxy indicators – Detailed on-site assessments
 - * QRS programs that have gone statewide are using complex measures

Kathy Thornburg

- We are working on planning with the four midwestern states (Kansas, Missouri, Iowa, and Nebraska) plus one outside the region (Mississippi).
- Each state has a rural and urban sample for its pilot.
- Our survey includes both programs that are receiving and not receiving subsidy.
- Each state QRS has 5 stars.
- States work together but use their own context.
- Issues:
 - * Funding
 - * Linking to state subsidy
 - * Documentation of education and training
 - * Number of classrooms per program to assess (Missouri did a study of validity on the percentage of classrooms used in their assessments)
 - * Using licensing as a base (what about unregulated care?)
- Missouri just adopted a statewide system based on their pilot study.

Mary Roach

- Wisconsin was planning an evaluation but did not move forward.
- Wisconsin did not recommend the use of ECERS – we are looking at another way to measure quality that is not as costly.

Matthew Richardson 10/30/06 10:30 AM
Deleted: 10

- The QRS proposal went down at the legislative level; QRS was not seen as a bad idea but the proposal needs more work.
- “Grow in Quality” Project (pilot instrument being used in 70 centers in 2006):
 - * Developed instrument that will measure quality
 - * Is using materials in classrooms
 - * Will try instrument out in the next year
 - * Hopes to determine what the most the important indicators are

Jamie Gottesman

- Step Up To Quality (SUTQ) is Ohio’s quality rating system:
 - * We’ve “gone from grass to mass quickly” (in little less than a year).
 - * The first center’s rating was just renewed.
 - * Stakeholders and key funders developed SUTQ about 8 years ago but it was then shelved due to funding problems.
 - * In 2003 work began again due to about \$10 million in unused quality dollars that needed to be expended.
- We have the resources to work with 1,000 centers (which would be 25% of all licensed center-based programs) across 9 counties.
 - * Pilot counties were chosen if they had strong local initiatives.
 - * The goal was always to include FCC and public preschool, but we had no resources or capacity to do that at the time of pilot implementation.
- Key considerations in Ohio:
 - * QRS benchmarks or components are fairly common among states based on research but there is very little in research that is definitive on what indicators should be chosen to best meet the benchmark.
 - * In other words how do you know that the indicators for each level for every benchmark actually “mean something?”
 - * We are currently evaluating if the way in which Ohio bundled the indicators together for each benchmark for each step buys a level of

Matthew Richardson 10/30/06 10:30 AM
Deleted: 10

quality as measured by the ECERS-R. (278 classrooms have been assessed.)

- * SUTQ demonstrates through the indicators related to staff qualifications and ongoing professional development that the rubber hits the road with what happens between child and teacher.
- * There is one essential indicator that is hard to operationalize and that, of course, is the disposition of the adults who teach children.
- Critical variables:
 - * How to get at the essential without being exhaustive?
 - * Observable not inferable – how do you know, what do you take as proof?
 - * Use data (state and national) to drive QRS decision-making.

Current Findings:

- * Step 1 overall score is 4.3
- * Step 2 overall score is 4.96
- * Step 3 overall score is 4.58
 - (This is the level with an associate of arts degree in every classroom, either accredited or with ratios lower than NAEYC, and 22 hours of annual professional development).
 - The personal care scores were the lowest in Step 3's and could be the reason for the lower scores. It begs the question though: is there a health/safety tradeoff when it comes to school readiness?

Questions and comments

Dave: How are you using ECERS (if it is not used to document a level now)?

Kathy: We're using it to validate bundles of indicators at each step (tie learning environment indicators to early learning guidelines)

Dave: Where does planning sit on a continuum of quality assessment?

Kathy: all 5 states in our project are using learning environment assessment, but not necessarily for all classrooms. We're leaning toward on-site observation but something not as costly.

Question: Curriculum for preschools – they tend to do well on ECERS-R. What to do for infant curriculum?

Mary: We are leaning toward proxy indicators. We have a lot of administrative data that we want to rely on. Directors can get more points

Matthew Richardson 10/30/06 10:30 AM
Deleted: 10

if they have an administrator credential – there is increased interest in earning it.

Question: What about other data: accreditation, food program participation?

Kathy: There's an article in our professional development book about food program data.

Question: How has each project looked at ECERS and how was it used? What was the training on ECERS? How are you maintaining reliability?

Kathy: We're concerned about this. We have a "gold standard" person in the four states. We have one person who goes to North Carolina and does training in multiple states, making sure there is cross-state reliability. In Missouri, every 10th observation has an inter-rater reliability check.

Mary: We are not using this instrument, though we have done many research observations using it in about 700 center classrooms and 70 family child care homes using FDCRS (Family Day Care Rating Scale), with training from North Carolina and an "anchor" person. We did not have a positive experience measuring quality with ECERS. We believe it is a blunt research instrument to put into QRS.

Jamie: We are committed to doing it right. Assessors are trained in North Carolina's method, and we have an anchor system (anchors must score 90 percent or above to be anchors). Regarding continuum, we want to base evaluations on evidence: Have internal protocols to interview teachers on planning process, do environmental scans, watch what happens with child progress. The care environment may not be stable throughout the day.

Dave: Are you doing intensive research on the 1,000 in your pilot project? Are you planning to use ECERS when you go to scale?

Jamie: There are many ways to use the scales when SUTQ goes statewide. One way might be to have an alternative pathway to reach the next step that uses overall rating scores in lieu of ratios. This thinking comes from our data that shows lower ratios are not yielding higher ECERS-R scores. Ratios seem to matter more in the absence of certain things and matter less in the presence of certain things.

Question: In Connecticut, we have a high percentage of accredited programs, and we have invested a lot in NAEYC accreditation (have 250 centers up for reaccreditation next year). Accreditation should be the top level in QRS. There is a debate on whether to do ECERS as a measure of steps toward accreditation. Why aren't more states using accreditation as the top level?

Matthew Richardson 10/30/06 10:30 AM
Deleted: 10

(From the floor): Rhode Island is grappling with accreditation. We're trying to develop a system for FCC, centers, and SAC – these have different accreditations and the systems aren't equal. We want to make levels equivalent for all providers. There's an ongoing conversation about the old NAEYC system – mistrust as to whether it was a good measure of quality. We don't know if can trust the new system to be a “gold standard.”

Kathy: MO has state accreditation for the 5th star. Programs have to have at least 4 stars. We have criteria for accrediting bodies and have approved seven: FCC, NAEYC, SAC, and state accreditation are the major ones. There are differences in accreditation, but child care staff carry out a self-assessment. Programs are learning from this. They become a 5 if they meet accreditation.

Comment: Participants from other states are having similar discussions. Some programs that are accredited have licensing compliance issues.

Dave: Some advocates for accreditation say that indicators are based on a solid base of research. We should have assurance that the program is high quality. Some states are developing steps to lead to accreditation at the top tier.

(From the floor): Because the new NAEYC process is based on research, accreditation would be at the 5th star. The National Association of Family Child Care (NAFCC) is not as rigorous, and accredited programs will be at the 4th star.

Question: Are you collecting information from parents?

Kathy: We had parents answer questions. In new model, we're not doing that – the answers didn't add anything. We asked directors what they provided for parents and then asked parents what they thought the program provided for them. We got the same information. It is less costly to get information from directors. An enhanced family involvement component was developed in response.

Question: The focus here is on implementing in centers, but QRS is difficult to implement in FCC. How can we make sure an FCC can reach the highest level? Is it affecting access to child care for low-income families? How can we make sure to include all providers?

Kathy: We developed the FCC, center, and SAC model at the same time. Being in compliance with licensing is the first star. There is a pre-QRS

Matthew Richardson 10/30/06 10:30 AM
Deleted: 10

component – if unregulated, you could move higher if you become licensed. This helped engage other providers.

Mary: In Wisconsin, 70% are in centers; that's where bulk of kids are. We want to include FCC and SAC, but later.

Kathy: In Missouri, the bulk of programs are family child care homes. Rating systems are identical except in training and education components, and there is no administrator component in FCC. Feedback from technical assistance with FCC is easier. There are fewer staff to deal with, making systemic changes is easier, and there are fewer adults to change. But these providers serve fewer children.

Dave: The Wisconsin proposal included centers, licensed FCC, and certified FCC (unlicensed but certified to receive subsidy). It's a challenge to have equivalent standards. What is a reasonable marker? States struggle with how to phase in all types of care and ask what to start with. In our study most of the standards were equivalent across types, allowing us to make comparisons. In our workforce data on education, for example, we found that average education levels among family child care providers were lower than among center providers.

Question: Do you have ideas on how to determine if the investment is paying off for kids? If there is limited money, how do you decide to put in QRS as compared to raising reimbursement rates? Do we have the data to determine what's the best mechanism? Are we collecting these data and looking at depth of child outcomes?

(Response from the floor): We did raise reimbursement rates in Rhode Island, but it's hard to maintain that the programs are good quality if we don't have data. We need QRS to sustain raised rates. Should we put money into measuring quality or into improving quality? We can't do one without the other.

(Response from the floor): In North Carolina, the 5 star program saw the 1 stars move to at least 3 stars over time. Level 1 is simple and inexpensive, 5 is more expensive. Since stars are tied to reimbursement, there is an incentive to raise quality. Education levels are going up and we have seen definite improvement in quality. We have used ECERS but don't use a child measure.

(Response from the floor): In Iowa most children are in unregulated care. We need pre-post quality measures. We are using child outcome measures in kindergarten where most children are attending.

Matthew Richardson 10/30/06 10:30 AM
Deleted: 10

Jamie: Child outcomes are one of the easier tasks. Pilot programs are part of early learning initiatives. We will have child data in relationship to language and literacy, are hoping to find multiplier effects. Some states invested in pre-kindergarten without demonstrating a capacity to provide the right dosage of intervention. No matter how you monitor, it is an honor system where programs promise to do the right thing for children. Accreditation can be a different experience for providers. High quality is essential for low-income children. How can we pay more to programs that do more? Other benefits: demand for professional development is huge, and the rating system is driving it. For early learning, it has to be *approved* professional development.

Comment: Technical assistance is very expensive. One county sent a bonus payment to rated programs.

Question: Why did ECERS scores go down on step 3 in Ohio?

Jamie: We don't know. Did program structure influence the ratings? No. Did accreditation have an impact? Accredited centers are across the steps. What about characteristics of the children? All but a few of the step 3s are serving the most vulnerable children. A lot of it may be health; nothing else matters if children are sick. Ohio's step 2 providers changed the most. This reflects a density factor in teachers. It's a "feast or famine" model; we put a lot into one person, but we need more people with better qualifications.

Question: In Rhode Island we are designing a pilot and using ECERS as process evaluation for a level. What can we use to measure quality for the pilot? Is there an independent research tool?

Response: Look at chapter by Rich Lambert in new handbook on ECE research. The Early Language and Literacy Classroom Observation (ELLCO) can be used in preschool, language and literacy assessment.

Dave: Wrap up comments: This is a new and developing field – and very complex. We need new perspectives across states in order to move the agenda forward.

End of Session

Breakout session notes are brief summaries of issues, findings and ideas discussed by participants and do not necessarily reflect the views of the Child Care Bureau or other members of the Child Care Policy Research Consortium.

Matthew Richardson 10/30/06 10:30 AM
Deleted: 10