2012 CCPRC Annual Meeting
Workshop Summary A-4
October 24, 2012, 10:45 a.m.-12:15 p.m.

Building a Research Agenda on Effective Monitoring and Quality Assurance in Early Childhood Systems

Description

The goal of this session was to develop key research questions and strategies to examine monitoring in early childhood systems. Considerations and recommendations for building a research agenda were discussed. Presenters offered reflections on key elements of monitoring and quality assurance such as data collection strategies (observation, documentation, and verification), training and reliability procedures for raters and licensors, and options/incentives for improving compliance. Presenters drew on evidence from other fields as possible. The facilitator engaged the presenters and the audience in a discussion about research questions and methods for assessing the effectiveness of different elements of the system.

Facilitator

Nancy Margie, OPRE

Presenters

Bentley Ponder, Georgia Department of Early Care and Learning
Bridget Hatfield, University of Virginia
Kyle Snow, National Association for the Education of Young Children (NAEYC)

Scribe

Ladia Albertson-Junkans, Child Trends

1. **Documents in Session Folder**
   - "Key Indicator Methodology;" Rick Fiene, NARA and OCC
   - "Supporting States in Enhancing Reliability of QRIS Ratings-What we are learning from CLASS;" Bridget Hatfield
   - "NAEYC Accreditation Monitoring and Quality Assurance;" Kyle Snow, Stephanie Olmore and Susan Hedges

2. **Brief Summary of Presentations**
   - **Summary of Presentation #1:** Bentley Ponder
     o In Georgia, the Department of Early Care and Education (DECAL) is distinct from the Department of Education. DECAL oversees the Georgia pre-k program, licensing and monitoring, the federal nutrition program, and quality initiatives. Child care subsidies were added to DECAL in 2012 in preparation for the Race to the Top-Early Learning Challenge grant application.
     o A homegrown tool used to assess environmental quality was replaced with the CLASS in the late 2000's as part of an effort to improve inter-rater reliability. During the first year using the CLASS, 92% of all preK classrooms in the State were

observed at 96.1% and 92.2% reliability for center-based classrooms and family child care. Since then, one third of classrooms are observed annually.

- o Large variations in health and safety compliance determinations across licensing consultants (and regions) raised questions regarding the validity of the system. As a result, Georgia commissioned a study by Rick Fiene to review several years of licensing data to evaluate whether compliance determination is a valid indicator for program health and safety (as a foundation for QRIS participation). That study should be completed by the end of 2012. Maintaining confidentiality has been a challenge throughout the study.
- o Question: Is the classroom context defined in rating process? Each consultant does 4 cycles of 2 hours, within 30 minutes of program starting; there are 4 pages of clarifications for possible scenarios.

Summary of Presentation #2: Bridget Hatfield
- o Bridget discussed the importance of defining quality and what matters for families and the need to measure that reliably, fairly and equitability. Failure to do so could be a "train wreck" with the possibility of inaccurate decisions and lawsuits.
- o Studies on the reliability and validity of CLASS have revealed that there is more variance between classrooms than within classrooms in a given day, raters are biased, and the consequences of CLASS scores can be significant.
- o The possible consequences of biased data within the QRIS framework can be high stakes; e.g., reauthorization, loss of subsidy vouchers, lower subsidy rates, and elimination of classrooms.
- o Potential solutions to improving reliability:
  - Select consultants with less bias: not someone within the program being rated; researchers as opposed to teachers and administrators; and groups with child centered beliefs about children.
  - Training: minimum of 2 days, consider additional training; 80% within 1 (certification cut-point), consider using a more rigorous level; trainer quality and turn-over; what makes a trainer more effective and participants more engaged? (unknown at this point).
  - Ongoing support: frequency and intensity of calibration and analyzing data in real team to inform decisions.
  - Teams of raters rather than individuals: averages more accurate than single score; mix up the teams.
- o We do not know how many classrooms should be sampled within programs and how much of a reliability boost we get from inputs (each additional observer, day, etc.).
- o Questions from the audience:
  - Looking at QRIS in the child care world, is there any likelihood of States being able to afford what it would take to ensure reliability? That's why some States use the instruments for continuous quality improvement rather than scores. It depends on the nature of the QRIS system; if data is used for high stakes purposes, then it's going to be important to make sure the data used is not biased to the extent possible.
  - What about the use of seeding to detect bias (based on generalizability theory)?

- What about the toddler version of the CLASS? The toddler version was just officially released and the infant version is still in pilot. Data is being collected to analyze differences in quality dimensions and child outcomes.
- To what extent will Rick Fiene's study allow for valid comparisons on core items that may be useful to policymakers and ACF as relates to program monitoring? Fiene pulled licensing studies from 2008 and 2009 to see whether the right items are being used. Core rules are looked at each time a consultant visits a provider including during complaint investigations (which means that there will be more data about some programs than others).

- **Summary of Presentation #3:** Kyle Snow
  - NAEYC accreditation was developed in the early 80s and a reinvented version was implemented in 2006; standards have evolved over time. It is a challenge to achieve alignment between increasing knowledge about what is important and accreditation standards.
  - Accreditation is a system/process with four steps (self-study, becoming an applicant, becoming a candidate (gatekeeper), and meeting/maintaining standards). NAEYC accreditation is about meeting standards, not scores on a particular measure. There are 10 standards and each standard includes multiple topics; there are 417 program criteria total. To be accredited, 80% of these set criteria have to be met within each standard.
  - Self-study is vulnerable to selection bias related to prior knowledge of accreditation and motivation.
  - There are four types of data: program and classroom portfolios, family teaching staff survey, and classroom site visits. Every criteria informed by at least 2 sources of evidence
  - Assessor reliability: site visitors are assessors; they have a bachelor's degree with focus on child development. They undergo a 5-day training, have an annual reliability visit, attend an annual assessor symposium, and have a quarterly performance report (which identifies sources of error and provides feedback on performance).
  - NAEYC's business model is intended to ensure quality in a cost effective way. As of October 2012, there were 6,800 programs with NAEYC accreditation serving approximately 700,000 children.

3. **Brief Summary of Discussion**
   - Ivelisse indicated her interest in areas of new research and focus toward doing a better job of what goes into QRIS systems.
     - Given different sources of evidence used, how valid and reliable are the processes for obtaining these sources of evidence?
     - What can be learned from NAEYC process and data sources to apply to quality determination and monitoring?
     - With different types of programs (child care, Head Start, pre-k, etc.), how can we improve efficiency?
   - Kyle--what are the pros/cons between using accreditation as a framework for improving quality and choices versus using QRIS as a framework for this? What about accreditation

in relationship to QRIS? NAEYC has a long history and has data that could be mined (link accreditation data with ECERS scores?).

- o Bridget, workforce issues? We're not paying staff to undergo the accreditation process. Teacher compensation issues.
- o Georgia has put resources into accreditation, and while it doesn't automatically allow providers to move to the top level, providers get extra points for accreditation.
- o There are issues around different accreditation mechanisms with some being more rigorous than others. This raises issues around procurement and dictating which accreditation body providers can use.
- o Need to do more around analysis of the cost of quality. Should we be shifting the country away from a market-based system to a cost-based system? We don't currently have enough data about costs to improve finance and pay teachers better.

4. Summary of Key issues raised

- *Emerging findings that may be of particular interest to policy-makers and ACF?*
  - o Inter-rater reliability is sub-optimal
  - o There are things we can do to minimize CLASS observer bias.
- *Methodological issues including innovative methodologies that may help maximize resources available for research and evaluation?*
  - o Maintaining confidentiality when matching state level administrative data
  - o Strategies such as seeding to detect bias and analysis of core items in licensing.
- *Follow-up activities to address questions and gaps?*
  - o Research questions to consider:
    - What is the optimal number of classrooms to sample within programs to ensure equitability and cost effectiveness?
    - How much of a boost in reliability is achieved from adding each observer, day, etc?
  - o Studies that might be of interest:
    - Piloting in school-age child care
    - Enhanced resources for self-study
    - Streamlined, transparent assessment tool.