

2015 CCPRC Annual Meeting

File Naming Convention: C2: Recent Findings from QRIS Validation Studies 12.2.15

1. Descriptive Information

<p>Workshop C-2 (<i>Independence B-C</i>) <i>Recent Findings from QRIS Validation Studies</i></p> <p>Description A panel of researchers will share analyses and insights from current QRIS validation studies followed by a facilitated discussion exploring key similarities and differences in the findings across studies and the potential implications for future QRIS research and implementation.</p>	<p>Facilitator • Kelly Maxwell, Child Trends</p> <p>Presenters • Shannon Lipscomb, Oregon State University - Cascades • Kathryn Tout, Child Trends • Ying-Chun Lin, University of Wisconsin - Madison</p> <p>Discussant: N/A</p> <p>Scribe • Jennifer Cleveland, Child Trends</p>
--	---

2. Documents in **Session Folder** (Please list any electronic documents or web links used during the session.)

3. Brief Summary of Presentations

- **Context: All presenters are from INQUIRE, sponsored by OPRE**
 - QRIS has historically emphasized child care and FCC; recently has incorporated pre-k and head start
 - State driven
 - Voluntary
 - Intended to be a stair step to higher quality – nothing in between licensing and accreditation.
 - States like the in between stage of licensing and accreditation
 - There is no ONE QRIS. Differences exist. Limitations in generalizability.
 - What is validation? Zellman and Fiene definition (2012)
 - Talk about validation in the context of RTT-ELC –
 - Do the ratings differentiate quality in a meaningful way?
 - Does children’s development vary depending on rating level of the program children attend?
- **Summary of Presentation #1: Shannon Lipscomb, Assistant professor of family science, co-lead of Oregon’s validation study of Oregon’s study**
 - Initial findings
 - Oregon’s QRIS has 5 levels
 - Sample:
 - 1 licensed
 - 2 commitment to quality (validation study only enrolled a “confirmed” 2)
 - 3, 4, 5 are rated via portfolio system (observations only done at level 5)
 - 5 domains (multiple standards = 33)
 - Primarily a Block system
 - Levels 3-5 are considered “quality”
 - Example of level 3
 - Uses curricula in specific ways

- Qualifications for teachers and assistants
 - Measured by professional development registry
 - 50% of teachers need to be at step 7
 - Example of level 5
 - Curricula more specific and additional uses required
 - Higher standards of professional development
 - Addition of observations of teacher child interactions
- Sample for study 1
 - 312 programs (about half FCC)
 - Programs by age group
 - Programs by QRIS level
 - Pretty good distribution except for level 4.
 - Evaluation distribution mirrors state distribution of qris enrollment
- CLASS
 - Pre-K, Toddler, and Combined (mixed ages et al., 2011)
- Analyses to date
 - Internal consistency (among standards within QRIS)
 - Correlations and Cronbach's alphas (all are high)
 - Links between QRIS ratings and CLASS scores
 - One score per program for PK, Toddler, and/or combined
 - Average across classrooms/groups
 - Correlations and ANOVAs with post hoc tests
 - Focus on lower (1-2) vs higher (3-5)
- Fast track vs fully rated
- Summary of CLASS scores by QRIS level
 - Statistically significant differences, examining CLASS scores for PK, Toddler, and Combo
 - Classroom organization consistently high in centers
 - Emotional support consistently high in FCC
- Strongest evidence for differences by QRIS ratings in instructional support – linear representation of class scores and star level
- Fast tracking
 - Descriptively the programs look very similar
 - Fast tracking seems pretty reasonable
- Variability across classrooms within programs
 - A lot of the variability is within the program
 - 28% - 43% of the variance within programs
- Next steps for multilevel models
 - Provider characteristics and CLASS scores?
- How meaningful are the differences?
 - Higher rated programs reaching a benchmark
 - This might be meaningful for examining thresholds of quality.
 - The effect sizes ~~look bigger than we expected to~~ are substantial
- Study 2 is going on now.

- **Summary of Presentation #2: Kathryn Tout, Co-Area Director Early Childhood Development, Child Trends**
 - Minnesota Parent Aware
 - MN is RTT state, but the evaluation is funded is by Parent Aware for school readiness (PASR)
 - Parent Aware
 - First two levels are block
 - Levels three and four are points

- Class is used at three and four levels
- No observations used in FCC
- Accelerated pathway – accredited, head start, Pre-K; must meet curriculum and assessment indicators
- 2015 statewide rollout
 - 90% density of accelerated
 - 10% density of fully rated
 - Star 4 most common rating (70%)
 - Star 2 is most common among full rating
- Validation questions
 - Approached all fully rated programs to participate
 - A lot of accelerated programs participated in the evaluation
 - For so many of these studies these have been done in the early rollout of the QRIS system
- Rating level and observed quality?
- Rating pathway and full rating?
- Does observed global quality in center based programs differ by rating level and rating pathway?
 - There is no difference between star 3-4 and APR
 - There is a difference between those programs and lower star rated programs
- CLASS?
 - IS differences for 1-2 vs 3-4
 - No differences for ES and CO
- ECERS-E
 - Language star APR vs 1-2 and APR vs 3-4
 - Math
 - Individualizing instruction **APR** vs 1-2
- Program Type
 - Three and four score higher on IS than accredited and school readiness.
- FCCERS-R
 - No differences between lower and higher rated programs
 - No differences in ECERS-E either
- ECERS-E
 - Program differences
- Implications
 - ECERS-E/R results provide one source of initial support for the validity of the PA ratings
 - The evidence on the accelerated pathway is mixed. The differences do not indicate a full rating process would be beneficial to implement at this time.
 - It is important to consider the role of CLASS coaching in the positive IS findings for fully rated centers
 - No support was evident for the validity of the ratings for FCC.
 - FCC are the fastest growing group in parent aware
- Key messages
 - PA is still early in statewide implementation
 - Even though the quality is differentiated by the ratings, the magnitude of the differences is small. The overall quality across rating levels and across program types is lower than expected.
 - Major changes not recommended at this time
 - Proposed changes should be approached cautiously

- **Summary of Presentation #3: Ying-Chun Lin, School of Social Work, U of WI, works with Katherine Magnuson**

- Wisconsin QRIS YoungStar hybrid system

- Operates more like a block than point system
- Programs are allowed to choose different rating pathways
 - Automated, Technical, or Formal
 - 4 domains
- Examples of requirements at different rating levels
 - 2 star – licensing requirements
- Young Star is more mandatory than other QRIS in the nation
- 17% of all programs in WI participate however pre k and school based not included
- 48% of family based enrolled
- Allow programs to participate but don't have to go through the rating process (these programs receive a 2 star automatically)
- Provide tiered reimbursement at different levels
- Goal of validation study: do the points make sense and work as intended?
 - Do higher rated programs have higher levels of higher levels of observed quality?
 - Do children have greater gains if attending higher rated programs?
- Northeastern region of WI and Milwaukee area chosen for sampling (high areas of higher income families and lower income families respectively)
- Number of classrooms observed by provider type and star rating
 - Majority of programs only have one classrooms observed
- How well does YoungStar rating predict ERS scores?
 - Significant difference in ERS score at 2 star vs 3-5 star
 - Between 3-5 star, there is not a statistical significant difference between each star level
- Examine each subscale of ERS
 - Some statistically significant differences detected
- Star Rating points (only programs with technical or formal rating)
 - Total YoungStar points significantly predict ERS scores
 - Looking within rating domain all sig predict ERS scores
- Summary and Conclusion
 - Both star level and points predict observed quality
 - Differences between 2 star and more highly rated programs
 - Nearly all the variation in the moderate quality range
 - ~~Young star points appear to be valid~~
 - Child outcomes to come soon
 - Validation is only one part of the larger issue regarding whether policy innovation is working as intended
 - Engaging providers who are not involved in receiving a formal rating or technical assistance
 - Issues of cost are important
 - State partners are eager for empirical evidence that can be used for improving the rating scale (without increasing costs)

4. Brief Summary of Discussion

- a. OR: Effect size: level 2 vs level 3-5
- b. WI: majority of sites only included one classroom per program
 - i. There is variation among providers but it's not as big as we expected
- c. Took about 3 years from when implementation started
- d. ECERS-R/E is low. CO and ES are high. Those are 6s for all types. The other source of evidence is the low scores of the ECERS-e/r and IS.
- e. Who is doing the ratings?
 - i. Independent observers
- f. Validation study isn't funded by RTT in MN. But, in WI, RTT funds the study. Do you have any guidance for them in terms of what they should be able to sustain?

- i. Not sure. The cost is really a concern.
 - ii. What about when RTT funding ends? MN will still have the rating system will remain the same. But the improvement supports may be affected. How do you keep the incentives to improve?
- g. Oregon: one of our additional side questions is what would happen if we had a more streamlined approach? Is there a way to use our findings for what is most helpful?
- h. Have you had a chance to look at the relationship of specific domains of the QRIS and quality?
 - i. OR, yes, can't talk
 - ii. MN, yes, constrained by sample size. We don't have as much data that have gone through the full rating. Again, let's look again in 3 years. We have so few programs still.
 - iii. WI, yes, professional practices has largest predictability when holding other domains constant. We suspect that we have a lot of programs that are chain private programs and these programs may have more resources. So when they are able to do certain business practices, they are also able to provide curriculum for example.
 - 1. Curriculum predictive but to a lesser extent.
- i. Comment
 - i. As much variability among classrooms. Could also be unreliability of the instrument itself.
 - ii. Lynn Karoly has an article in ECRQ.
- j. When will reports be available?
 - i. WI – public. WI YoungStar website, exec summary is there and full report is there
 - ii. MN – January www.pasrmn.org, Research Connections
 - iii. OR email Shannon in 2016