Oregon's Quality Rating Improvement System (QRIS) Validation Study One: Associations with Observed Program Quality

EXECUTIVE SUMMARY

Oregon State
UNIVERSITY **OSU**

Shannon T. Lipscomb, Ph.D.,
Associate Professor of Human Development and Family Science,
Oregon State University-Cascades

Roberta B. Weber, Ph.D.,
Faculty Research Associate, College of Public Health and Human Science, Oregon State University

Beth L. Green, Ph.D.,
Director of Early Childhood and Family Support Research, Portland State University

Lindsey B. Patterson, M.S.,
Senior Research Assistant, Center for Child and Family Studies, Portland State University

Portland
State
UNIVERSITY

Acknowledgements

Oregon's QRIS Validation Study One

Executive Summary

Introduction

Nationally, Quality Rating and Improvement Systems (QRIS) have emerged from concerns that large portions of American early care and education (ECE) programs were not of high enough quality to support children's development (Helburn, 1995; NICHD, 2003). As of 2015 all states were planning, piloting, or fully implementing a QRIS (Build, 2015).

Oregon's QRIS is a comprehensive system composed of standards, supports, incentives, consumer education, and rating/monitoring. All types of regulated providers in all parts of the state are encouraged to become rated. Standards are clustered into five domains: learning and development, personnel qualifications, family partnerships, health and safety, and administration and business practices.

Oregon has mostly a "building blocks" system, which means that programs must pass all or most of the standards for the 3-, 4-, or 5-star level to achieve a rating at that level. Level 1 of Oregon's QRIS represents programs that are licensed but have not voluntarily participated in the rating process. Level 2 (termed "Commitment to Quality" or "C2Q") indicates that the program has made a formal commitment to quality improvement by attending a QRIS training. Many of these Level 2 programs have not submitted portfolios; others have submitted a portfolio but did not earn a rating of 3 or higher. Programs are only required to submit materials specifically related to the star level for which they are applying. Accredited and Head Start programs only needed to submit documentation on standards not included in NAEYC or Head Start/Early Head Start standards. The QRIS ratings also rely on data from licensing and the Oregon Registry Online.

This Validation Study
The study described in this report is the first of two studies on the validity of Oregon's QRIS. This study uses a measure of the observed quality of adult-child interactions as a benchmark against which to compare QRIS ratings.

Research Questions
1. What is the quality of programs in the QRIS Validation Study, as indicated by CLASS scores and QRIS ratings?
2. How highly correlated are the QRIS domains and standards with one another?
3. How well do programs' QRIS ratings differentiate observed quality of adult-child interactions?
4. How do certain QRIS standards & indicators of interest relate to observed quality?
5. How well are other personnel measures associated with observed quality and final QRIS ratings?

Methods

Sample
The Validation Study sample included 304 programs (levels 1-5) that were observed using standardized measures of adult-child interaction quality. Some analyses were only possible to conduct with a subsample of programs (N = 246) that had QRIS rating data (levels 2-5). Level 1 programs were

identified through other existing data sources but did not submit portfolios to the QRIS system.   The sample represented all three types of regulated programs in Oregon: 65 (21%) Registered Family (RF); 94 (30%) Certified Family (CF); and 153 (49%) Certified Centers. Observed programs served children between the ages of 15 to 60 months (i.e., toddlers and preschoolers). Programs in the sample ranged in size from those with only a single group/classroom to centers with up to 25 classrooms.

Measures

QRIS Ratings. QRIS ratings included 3-, 4-, and 5-star ratings, as well as Level 2 programs that applied for but did not achieve a 3-star rating. Programs at Level 1 were a) licensed, b) not otherwise participating in the QRIS, and c) identified by Structural Indicator data to be unlikely to meet QRIS standards. Data included overall star ratings, domain scores, and ratings for each of the specific standards of the QRIS. The QRIS provided ratings for standards within five domains: (1) Learning & Development; (2) Personnel Qualifications; (3) Administration & Business Practices; (4) Health & Safety; and (5) Family Partnerships.

Classroom Assessment Scoring System (CLASS). Observations of adult-child interactions were conducted using the Toddler (15-36 months) and PreK (36-60 months) CLASS tools (see La Paro, Hamre, & Pianta, 2012 and Pianta, La Paro, & Hamre, 2008, respectively). For classrooms/groups with a mix of toddlers and preschoolers, a third tool ("Combined CLASS") was used (Joseph, Feldmen, Brennan, Naslund, Phillips, & Petras, 2011). The CLASS yielded scores on three aspects of quality: Instructional Support, Organizational Support, and Emotional Support.

Observations were conducted in up to 4 randomly selected classrooms within each program. CLASS scores range from 1 (very low) to 7 (very high). Ratings of 1 or 2 are "low range," 3 to 5 are "mid-range," and 6 to 7 are "high range". Scores were averaged across classrooms/groups for each program.

Oregon Registry Online (ORO) Data. ORO Registry Online is a statewide database of training, education, and demographics for persons employed in child care and education.

Structural Indicators (SI) of Quality. SI are measured for all regulated facilities in Oregon at the time of licensing renewal and include: teacher education, teacher training, teacher retention, teacher compensation—wages and benefits, and accreditation.

Results

Question 1) What is the quality of programs in the QRIS Validation Study, indicated by CLASS scores and QRIS ratings?

QRIS Ratings. Of the 246 programs with QRIS ratings (2 through 5), over one-third (37%) were Level 2, nearly one-third were star-Level 3 (30%) and one-third were rated star-Levels 4 or 5 (33%). A much lower percent (13%) of the Registered Family providers reached star-levels 4 or 5, compared with Certified Family programs (40%), and Certified Centers (36%). Certain standards were much harder for programs than others, especially LD9 (screening & assessment), LD11 (adult-child interactions), and HS6 (screen time). For Registered Family providers LD1 (philosophy), LD7 (planned activities), HS1 (health/hygiene instruction), HS3 (healthy eating), PQ1 (leader qualifications), and AB5 (program evaluation) were also very difficult.

CLASS Scores. Overall program-level average CLASS scores in the Validation Study were in the upper end of the "mid" range for Emotional Support (approximately 5.0) and Organizational Support (4.5), and at the upper end of the "low" range for Instructional Support (2.5). These scores are similar to those documented in other studies using the CLASS (Hatfield et al., 2016; Burchinal et al., 2010).

> Overall, Registered Family programs provided similar levels of quality in observed adult-child interactions as Centers and Certified Family programs.
>
> Yet, their QRIS ratings tend to be lower.

Registered Family programs provided similar levels of quality in observed adult-child interactions as Centers and Certified Family programs. Yet, their QRIS ratings tended to be lower.

Question 2) How highly correlated are the QRIS domains and standards with one another?

The five domains of the QRIS were highly correlated, as were the standards within each domain. This was likely a result of the portfolio/block structure of Oregon's QRIS. These high correlations present three primary challenges:

1) QRIS rating data do not appear to be capturing the full variability of programs' actual practices in each of the five domains, and/or differences between programs practices across different domains (e.g. Learning and Development versus Family Partnerships).
2) It is very difficult to identify specific standards and/or domains of the QRIS that are most clearly linked with observed quality. The correlation between a given standard and observed quality reflects not only the actual association among the standard and observed quality, but also the links between other standards and observed quality.
3) High inter-correlations mean that individual standards and/or domains do not contribute much unique or additional information about programs.

Potential solutions include a) changing the structure to a hybrid or points-based system that captures more of the natural variation in programs' strengths and limitations, b) reducing the number of standards and/or domains to reduce redundancies, and/or c) increasing the use of personnel measures that the study found best able to capture personnel qualifications and training.

Question 3) How well do programs' QRIS ratings differentiate observed quality of adult-child interactions?

Overall, programs that achieved a 3-, 4-, or 5- star rating had significantly higher quality adult-child interactions, as measured by the CLASS, than those at level 1 or 2. These differences were small to medium in size, depending on type of program and the age group of children or CLASS tool examined. Differences in CLASS scores were most consistently related to lower observed quality in Level 1 programs; differences were smaller and less consistent when only comparing programs rated 2 versus 3-star or higher. Results did not detect differences in observed quality between programs rated 1 vs 2, or between programs rated 3 vs 4 or 5, or between programs rated 5 vs those rated 3 or 4.

As shown in Figure 1, the vast majority of the differences in observed quality by QRIS ratings were for the Instructional Domain of the CLASS. Fewer differences were detected for the Organizational domains, and almost none were detected for the Emotional domain.

4

Figure 1. CLASS scores by QRIS rating for each program type
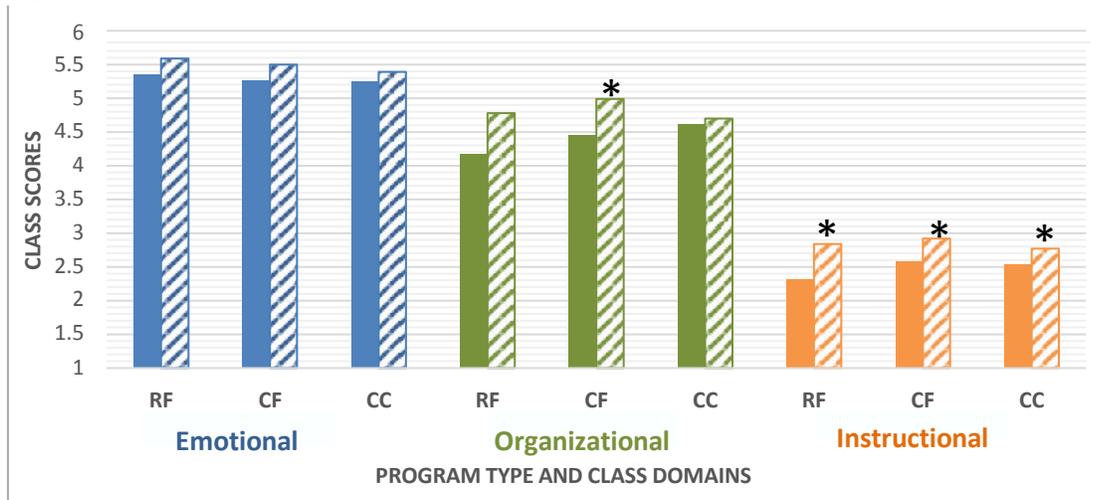
* Differences between programs rated 1-2 and 3-5 are statistically significant.
Program types are: Registered Family (RF), Certified Family (CF), Certified Center (CC).
Programs' CLASS scores represent an average across the PreK, Toddler, and/or Combined CLASS.

There are several possible reasons that links between QRIS ratings and CLASS scores were not larger:

- Many programs with high quality adult-child interactions were not successful in achieving a 3-star rating or higher. Twenty to thirty percent of the programs rated a 2 on Oregon's QRIS had among the highest CLASS scores in the study.
- The quality of adult-child interactions varied substantially by classroom/group within programs. This limited the strength of associations between programs' QRIS ratings and observed quality.
- The differences between higher- and lower-quality programs were small. For example, Instructional Support scores ranged from around 2.2 (for programs rated 1 or 2) to around 2.8 (for programs rated 4 or 5) on a scale from 1 to 7. These differences simply were not large enough to translate into large associations between QRIS ratings and observed quality.

> Programs that achieved a 3-star rating or higher on the QRIS showed higher-quality adult-child interactions than those rated 1 or 2.
>
> Yet, findings do not provide evidence that programs rated 4- or 5-star provide higher quality care than those rated 3-star.

Question 4) How do specific QRIS standards & indicators of interest relate to observed quality?

Findings from exploratory analysis of specific QRIS standards revealed some small, significant links between specific standards and observed quality on the CLASS. Given the high correlations among the QRIS standards we are more confident in identifying standards that are *not* well-linked with the CLASS than we are in identifying "the few and powerful" QRIS standards.

> Findings from this exploratory analysis revealed some small, significant links between specific standards and observed quality.
>
> Yet, concerns about several standards that were not linked with observed quality were also identified.

Many of the standards were either not linked with the CLASS, or were only inconsistently linked with the CLASS (e.g. for a specific CLASS domain, program type, or CLASS tool). This was particularly the case for the Emotional and Organizational domains of the CLASS; more standards were linked with the Instructional domain. Fewer standards were associated with CLASS scores for Registered Family programs.

These concerns may be important to consider, alongside other sources of information, in efforts to strengthen Oregon's QRIS. Findings revealed substantial concerns regarding LD9 (screening & assessment), 11 (adult-child interaction), and 12 (social and emotional development); we suggest either eliminating or substantially revising these standards. Additional standards that should be considered as candidates for elimination or revision include LD1, 4, and 6. Additionally, the Validation team found that the LD domain could be strengthened by combining LD2 and LD7 into one new standard.

Question 5) **How well are other personnel measures associated with observed quality and QRIS ratings?**

By accessing two additional sets of personnel measures from Oregon Registry Online that were not part of QRIS ratings the Validation Study team was able to more adequately assess the associations of personnel measures with observed quality.

For Centers, the personnel measures most closely linked with observed quality were: director registry step, teachers having either step 9 or higher, or a degree, and the median step for assistants. For Certified Family programs, the personnel measures most well-linked with observed quality were the provider's step or degree, assistants having a step 5 or higher, and staff training hours. For Registered Family programs, the only personnel measure clearly linked with observed quality was staff training. The associations between the providers'

> Personnel measures constructed from ORO, such as the Structural Indicators, were at least as consistently linked with CLASS scores as were the PQ ratings.
>
> This increases confidence in validation findings and points to ORO as an efficient source of personnel data linked to quality.

registry step and the CLASS were suggestive of a possible relationship but were not statistically significant, likely due to limited power from a small sample size.

Slight variations in how variables were constructed from the ORO database often led to differences in their associations with observed quality. Careful attention must be paid to how to utilize the ORO data.

Additionally, personnel measures, especially training, in Centers appear more complicated than for Family programs, possibly due to the larger numbers of personnel in centers.

There was a fairly strong link between the qualifications and training of the personnel in a program and the final star rating that program achieves. Finally, evidence that the Structural Indicator measures of personnel are correlated with both CLASS scores and QRIS final star ratings increases confidence in Oregon's ability to provide meaningful data related to the quality of programs that do not participate in the voluntary rating portion of QRIS.

> The Structural Indicators provide meaningful data related to the quality of all regulated programs in Oregon, including those not participating in the QRIS.

Considerations and Implications for Oregon's QRIS

Are Differences in Quality Sufficient?

Findings from this first validation study of Oregon's QRIS suggest that the QRIS somewhat differentiates the quality of the interactions that young children have with the adults that care for them in regulated programs. Yet, differences tended to be small, and only apparent when contrasting programs rated 3-star or higher to those at level 1 or 2. We did not find evidence that programs rated 4- or 5-star provided higher quality care than those rated 3-star. If Oregon's QRIS truly intends for 4- and/or 5-star ratings to represent higher quality care for children the rating system will need to be strengthened.

> If Oregon's QRIS truly intends for 4- and/or 5-star ratings to represent higher quality care for children than 3-star the rating system will need to be strengthened.

Most of the differences in observed quality by QRIS ratings were for the Instructional Support domain. QRIS ratings for Certified Family programs on the Organizational domain were also detected. Young children who receive higher quality care, especially in Instructional Support, show stronger school readiness (e.g. Hamre, Hatfield, Pianta, & Jamil, 2014). How much of a difference in quality is enough to improve child outcomes, however, remains unclear (e.g. Burchinal et al., 2010; Hatfield et al., 2016). In other words, children attending programs rated 3-star or higher appear to experience somewhat higher quality interactions with their teachers/caregivers than those attending level 1 or 2 programs, but whether this difference is large enough to translate into better outcomes for children remains unknown. Findings from studies of other QRISs across the country are mixed (e.g. Karoly, 2014). Study Two of Oregon's QRIS Validation Study is currently examining links between QRIS ratings and measures of child and family engagement.

> Whether the differences in quality between programs rated 3-stars or higher and level 1 and 2 programs are large enough to translate into better outcomes for children remains unknown.

Does Oregon's QRIS Represent Quality for all Types of Regulated Programs?

The conclusion, that programs rated 3-star or higher provide somewhat higher quality care than level 1 and 2 programs, is consistent across all three types of programs. However, although Registered Family programs provided similar levels of quality care to children as Centers and Certified Family programs,

their QRIS ratings tended to be lower. Few achieved 4- or 5-star ratings. This discrepancy calls for revisions to Oregon's QRIS to better reflect quality of care provided by Registered Family providers.

Additionally, findings highlighted the challenges of using a program-level rating to represent the experience of children in individual classrooms. Observed quality varied substantially across classrooms/groups within programs. Presently, Oregon's QRIS allows for such variability, such as by requiring group size/ratio patterns for only one age group, and/or by specifying that a percentage of personnel must reach certain qualifications. Findings indicated that this type of variation in teachers' and assistants' qualifications and training made it difficult to measure personnel qualifications in Centers, and to link them with observed quality.

Which QRIS Standards Work Best? Due to the primarily block-type structure of Oregon's QRIS, we have the most confidence in the validation findings for the overall ratings. Exploratory analyses that focused on the Learning and Development and Personnel Qualifications domains provided insights regarding specific standards, but revealed more about standards that were concerning than about standards best linked with observed quality.

Findings revealed substantial concerns regarding LD9 (screening and assessment), LD11 (adult-child interactions), and LD12 (social and emotional development), as well as some concerns regarding LD1 (philosophy), LD4 (indoor furnishings), and LD6 (materials). Additionally, the Validation Study team found that combining LD2 (curriculum) and LD7 (planned activities) into one new standard could strengthen the LD domain. We also have confidence that personnel qualifications and/or training are linked with observed quality, due to triangulating evidence across multiple sources of data.

Considerations for QRIS revision
- If a goal of the QRIS is that 4- and 5-star programs provide higher quality care to children than 3-star programs the ratings must be strengthened.
- Revisions should be made to reduce barriers to achieving 4- and 5-star ratings for the Registered Family programs that provide higher quality care to children equivalent to those in 4- and 5-star Centers and Certified Family programs (the bullets below provide concrete ideas).
- Consider changing the rating structure to a hybrid or points-based system that captures more of the natural variation in programs' strengths and limitations.
- Eliminate or substantially revise LD9, 11, and 12.
- Consider eliminating or revising LD1, 4, and 6.
- Combine LD2 and LD7 into one new standard, as described in this Validation Study.
- Streamline other standards and domains that are less directly linked with observed quality; the current study focused on LD and PQ because of theoretical links with observed quality.
- Consider increasing consistency in requirements across classrooms/groups in programs with more than one classroom/group. This increased rigor could be offset by eliminating standards that create barriers to achieving ratings without relating to observed quality.
- Consider other personnel measures from ORO, as possible replacements for the current PQ standards, and as supplemental data related to quality for all regulated programs in Oregon. Ensure that personnel measures remain intuitive and understandable to providers.