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

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

Acknowledgements

This work was funded in part by the Early Learning Division, Oregon Department of Education, as a part of its Race to the Top Early Learning Challenge Initiative. The contents are solely the responsibility of the authors and do not represent the official views of the funding agencies, nor does publication in any way constitute an endorsement by the funding agency.

Special thanks go to members of the Validation Study Mini-Review Team whose careful reviews strengthened this work. Team members included representatives of the Early Learning Division, The Research Institute at Western Oregon University, the Oregon Center for Career Development, and a Quality Improvement Specialist from a Child Care Resource and Referral agency.

We are especially grateful for the participation of child care providers who opened their homes and centers to our research team, and to the research assistants who collected data across the state to make this study possible.

Table of Contents

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.

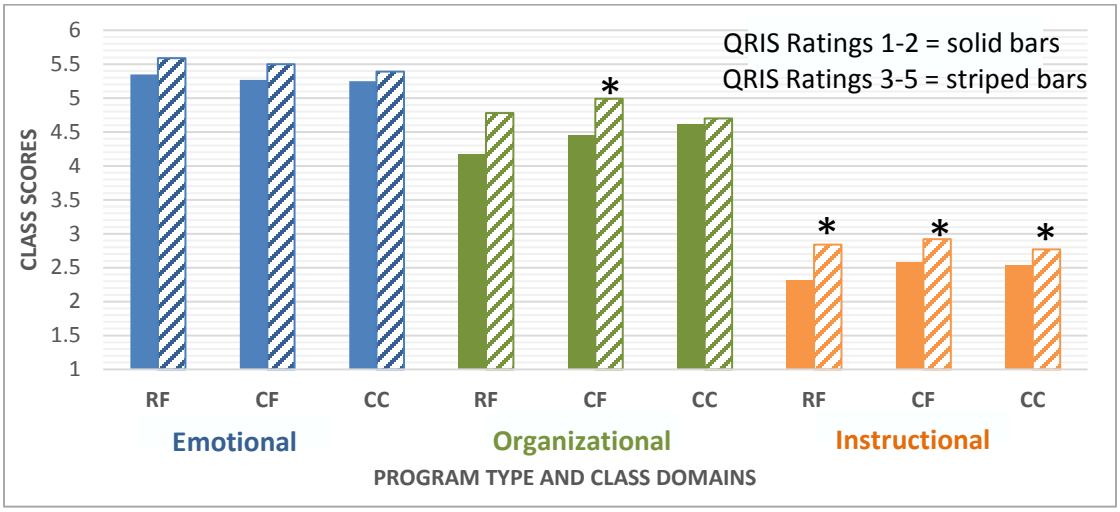
Potential solutions include a) changing the structure to a hybrid or points-based system that captures more of the natural variation in programs' strengths and limitations, b) reducing the number of standards and/or domains to reduce redundancies, and/or c) increasing the use of personnel measures that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

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.

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.

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' 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.

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

Introduction

National Context

Nationally, Quality Rating and Improvement Systems (QRIS) have emerged from the convergence of multiple concerns about the well-being of children. A number of studies have shown 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). Increases in parental use of ECE (Laughlin, 2013) and growth in the body of literature connecting quality to child outcomes led to concern that ECE programs were of insufficient quality to support school readiness and other measures of child well-being. States responded by creating quality improvement initiatives. One such strategy, used first in Oklahoma in 1998, brought quality initiatives together in a systematic approach that became known as Quality Rating and Improvement Systems. The Department of Education's Race to the Top (RTT) Early Learning Challenge grants' focus on QRIS heightened awareness of it and brought funding to build the system in states, such as Oregon, that received a RTT grant. As of 2015 all states were planning, piloting, or fully implementing a QRIS (Build, 2015).

QRIS Overview

Although each QRIS is unique, they share core features and functions. They involve both a rating component and an improvement component. Ratings are typically based on a set of domains or standards that are scored and then used to create an overall program rating (e.g. on a scale from 1 to 5). These ratings then set the foundation toward which quality improvements are targeted. In this way, a QRIS is a framework or system upon which quality improvement efforts are built; a QRIS is not an intervention. Some QRIS systems are voluntary, others are required as part of licensing (Child Trends and Build Initiative, 2016).

States have built their own QRIS systems, but along the way have had easy access to information on how other states constructed their QRIS through the *QRIS Compendium* (Child Trends and Build Initiative, 2016). States have also received technical assistance from organizations such as the Build Initiative and the QRIS National Learning Network. In building their systems, states have used the body of research that identified characteristics of ECE programs that are associated with positive child outcomes (e.g. qualifications of personnel, quality of adult-child interactions, use of assessment to guide instruction). In addition, states often create standards in areas such as health and safety or business practices that stakeholders believe are essential to being a high quality program.

Oregon's QRIS

Oregon's Quality Rating and Improvement System (QRIS) grew out of a long-term commitment to improving the quality of early learning in Oregon. In the mid-2000s a public-private partnership known as the Education and Quality Investment Partnership (EQUIP) built upon and expanded quality improvement initiatives already in place. These included providing scholarships and incentives for increased education and training to members of the early learning workforce. Two predecessors to the rating portion of QRIS emerged out of this effort: Quality Indicators and Oregon Programs of Quality.

Quality Indicators involved measuring all regulated programs on a set of indicators that research indicated were associated with positive child outcomes. That initiative has continued and is now known as the Structural Indicator project. Oregon Programs of Quality (OPQ) involved recruiting programs committed to improving quality, supporting these efforts, and awarding the Oregon Program of Quality designation to those that met an established standard of quality. OPQ proved to be a testing ground for the portfolio measurement system that was later adopted when Oregon developed a QRIS.

Oregon's QRIS is a comprehensive system composed of standards, supports, incentives, consumer education, and rating/monitoring. Although QRIS involves all regulated early learning programs, ratings are voluntary. 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. The full list of 33 standards is included as Appendix A. Oregon has mostly a "building blocks" system, which means that programs must pass all or most of the standards for the 3-star Level to achieve a 3-star rating. Some states use points-based systems in which programs need to earn a certain number of points in various categories to achieve a star rating. Points-based and hybrid systems tend to be more flexible than block systems in how programs achieve ratings (Child Trends and Build Initiative, 2016). Oregon's QRIS does have some flexibility (e.g. programs need to meet only 10 out of the 12 standards within the learning and development domain in order to achieve a star-rating), but typically operates more like a block system; programs submit evidence for the standards set forth at the 3-star Level in order to achieve a 3-star rating; they are not assessed in terms of whether they might achieve higher levels of quality (i.e., 4-star, 5-star) unless they specifically submit materials for that higher level. For more information about Oregon's QRIS see http://triwou.org/projects/qris.

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. Thus, the level of quality provided in Level 1 and Level 2 programs is not *necessarily* lower than those rated at 3,4, or 5-star levels; rather, it is simply *unknown* as they have not gone through the QRIS review process. To achieve a rating of 3-, 4-, or 5-stars, programs submit portfolios documenting achievement of standards at the given star-Level. Programs are only required to submit materials specifically related to the star-Level for which they are applying. In addition to the information submitted in their portfolios, the QRIS ratings also rely on data from licensing and the Oregon Registry Online.

Oregon QRIS leaders created a cross-walk of QRIS standards with those used in National Association for the Education of Young Children (NAEYC) accreditation and Head Start/Early Head Start program monitoring. Based on that alignment, accredited and Head Start programs were fast-tracked and only needed to submit documentation on standards not included in NAEYC or Head Start/Early Head Start standards (see Appendix B for the list of the cross-walked standards). Field-testing of Oregon's QRIS began in selected areas of the state in early 2013 and went statewide shortly after that. The State intends to implement a revised QRIS system in 2017. The State is currently engaged in an extensive process of information gathering about the QRIS, of which this Validation Study is a critical piece.

QRIS Validation

Nationally, as the prevalence of QRIS systems has increased, the validity of the ratings has emerged as a major concern. Parents are encouraged to use QRIS ratings in making child care selections. In many states, eligibility for funding or the level of funding is tied to the rating level a program achieves. States make major investments in producing ratings. Thus, there has arisen a demand for research showing the extent to which higher QRIS ratings are associated with external measures of quality or with more positive child outcomes. Policy makers and funders want assurance that highly rated programs actually provide care that is better for children. In her 2014 review of QRIS validation studies, Karoly described 14 early studies that validated ratings against measures of program quality or child outcomes. She concluded that studies using independent measures of quality have not found consistently positive associations between ratings and observed quality, and that the few studies using child outcome measures have generally not found the expected gains. Karoly argued that early studies often had methodological issues that could explain the mixed and often weak findings and made the case for stronger validation study designs. The Race to the Top Early Learning Challenge grants have included funding for well-designed QRIS validation studies of which this is one.

This Validation Study

Oregon's QRIS Validation Study has two goals: 1) to examine how well the QRIS rating system differentiates levels of observed program quality and child/family engagement and 2) to identify revisions that could enhance validity. Oregon is conducting two studies to accomplish these goals. The first uses a standardized measure of the observed quality of adult-child interactions (the Classroom Assessment Scoring System, or CLASS), as a benchmark against which to compare QRIS ratings. Results of this first study are included in this report. The second study, currently underway, uses measures of child and family engagement as a way to assess the predictive value of QRIS ratings at the child and family level; data from the second study are not included in this report. It is important to note that the current study (Study 1) is not designed to answer the question about whether or not the QRIS "works" or is "effective," but rather to assess the extent to which QRIS ratings are consistent with other sources of information about program quality (namely, the quality of adult-child interactions).

In other words, the Oregon QRIS Validation Study reported in this document examines how well the ratings that programs earn in Oregon's QRIS represent the quality of children's experiences, measured by adult-child interactions. More specifically, we examine five inter-related research questions.

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

Study Design Overview

The QRIS Validation Study utilized a non-experimental design, integrating data from 5 different data sources. To adequately address the study aims it was important to include programs within each of the 5 QRIS levels even though Oregon's QRIS only fully rates programs at the 3-, 4-, and 5-star levels. Thus, this study included programs at Level 1 who were a) licensed; b) not otherwise participating in the QRIS; and c) identified by Structural Indicator data to be unlikely to meet QRIS standards. The study also included those Level 2 programs that applied for a 3-star rating or higher but did not achieve it. Finally, the study included programs rated at each of the 3-, 4-, and 5-star levels. We provide an overview of the sampling structure and data sources here and then describe the sample, measures, and procedures later in the Methods section.

Data Sources

Data for the first QRIS Validation Study were collected from 5 different sources: (1) QRIS Rating Data from The Research Institute (TRI) at Western Oregon University (WOU); (2) Oregon Registry Online data regarding child care director/owner and provider/teacher qualifications, housed in the Oregon Center for Career Development at Portland State University (PSU); (3) Classroom Assessment Scoring System (CLASS) observational data collected by the Center for Improvement of Child and Family Services at PSU; (4) child care provider/teacher survey data collected by PSU; and (5) Structural Indicators of Quality data from the Hallie Ford Center for Children and Families at Oregon State University (OSU). Refer to Appendix C for a graphical representation of the five sources of data.

1. QRIS Rating Data. QRIS rating data for those child care facilities that voluntarily submitted a portfolio to TRI as part of QRIS were sent directly to the data analysis team at OSU at the conclusion of the data collection phase. PSU staff who collected CLASS observations were blind to the QRIS ratings made by TRI.

2. Oregon Registry Online (ORO). TRI retrieved child care director/owner and provider/teacher qualifications, including education, professional certifications, and ORO steps, from the Oregon Center for Career Development at PSU at the time the portfolio was processed. These data were sent to the data management team at OSU. The observational data collection team within the Center for Improvement of Child and Family Services at PSU did not have access to child care director/owner and provider/teacher qualification data provided by ORO.

3. Classroom Assessment Scoring System (CLASS) Observations. Observational data on adult-child interactions were collected by the Center for Improvement of Child and Family Services at PSU and sent directly to the data analysis team at OSU at the conclusion of the data collection period. Observational data were not shared with TRI at WOU and did not impact QRIS ratings. CLASS observations were conducted between July 2013 and July 2015.

4. Child Care Provider/Teacher Survey. As part of the QRIS Validation Study, surveys were collected from providers/teachers who were observed by the Center for Improvement of Child and Family Services at PSU. For the purposes of this study, data from the surveys are only used to describe the sample of teachers/providers who were observed.

5. Structural Indicators of Quality. The 2012 and 2014 Structural Indicators of quality databases were used for this study. These data were sent directly to the data management team at OSU. The 2013 structural indicators data were used to identify programs that were likely to not meet the QRIS standards. The resulting list of programs served as the pool of "Level 1" programs and was sent to PSU for recruitment into the QRIS Validation Study. The 2014 structural indicators data were used as an independent measure of quality. The 2014 dataset was used since the majority of portfolios were submitted in that year.

PSU entered and stored all CLASS observation and child care director and provider survey data during the course of the study. Similarly, TRI, Oregon Center for Career Development, and the OSU team with the Hallie Ford Center for Children and Families housed databases for their own data, separate from the other data sources. Once the data management team at OSU received databases from all five sources, they merged the databases and followed up with each data source independently for any emerging questions or issues. The final database, including information from all five data resources, was accessed solely by the team at OSU and was not shared with any of the originators of the original five data sources. Inconsistency in person-level identification numbers prevented a match of all provider staff across all databases. The data management team resolved some of the missing identification by working with the child care licensing staff at the Early Learning Division. The resulting missing data in analyses of ORO and PSU survey are noted in reports of results.

Procedures

Sample Identification: Programs Participating in the QRIS. The Center for Improvement of Child and Family Services at PSU had their plan for protecting human subjects approved by the PSU Institutional Review Board (IRB) prior to data collection in 2013 and OSU IRB accepted that plan approval. During the first 15 months of QRIS implementation (July 2013 through October 2014), TRI sent a list of programs that had submitted a QRIS portfolio and their contact information (i.e., program name, license number, director name, phone number, and address) to the Center for Improvement of Child and Family Services at PSU on a bi-monthly basis. The vast majority of these programs eventually received their QRIS rating from TRI; this rating was given independent of ongoing data collection by the QRIS Validation Study team at PSU. PSU contacted each program to determine whether they were eligible for the Validation Study. Programs were considered eligible if they served children between 15 and 60 months, spoke English or Spanish in the classroom, and were not exclusively focused on "after school" care or preparing teenagers (i.e., minors) for a career in early childhood.

Sample Identification: Level 1 Programs. Additional child care programs that had not and were not planning to submit a QRIS portfolio to TRI in the next 6 months were identified by the data management team at OSU using structural indicator data which included information on personnel qualifications. These programs represented "low quality" programs (i.e., Level 1 programs) as suggested by the structural indicator data from the year 2012. Structural indicators varied somewhat by type of care; the study identified indicators that were fairly equivalent across Centers, Registered Family, and Certified Family. For Centers the criteria were that the program had a) 25% or fewer teachers at step 7 on the Registry and 25% or fewer teachers had "some college/degree" *in the field*; b) Director did not have a step 8 on the Registry and did not have at least "some college/degree" *in the field;* and c) fewer than 75% staff had 18+ hours of training in the past year.

To be eligible for the QRIS Validation Study, Level 1 programs also had to be similar to QRIS-participating programs by being in business for at least two years and not having any validated complaints. These Level 1 programs were then stratified by ages of children served, geographic location (metro versus non-metro), and type of care, and then randomly selected for recruitment. A list of these program names and their contact information were sent from OSU to the data collection team at PSU for QRIS Validation Study recruitment.

Program Recruitment. Child care programs identified through either of the two sample identification strategies just described were contacted by PSU staff and invited to participate in the QRIS Validation Study. The child care program contact information was then transferred to one of PSU's 12 data collectors across the State. These data collectors were blind to QRIS ratings. Additionally, data collectors were unaware that programs not participating in QRIS were identified as "lower quality" or "Level 1 programs" through the structural indicators data. Instead, PSU data collectors were told that the list of programs not participating in QRIS came from the State and were identified only because of their lack of participation in QRIS; thus, they could have any level of quality from low to high.

Data collectors then contacted the program director/owner by email and/or by phone to tell them about the QRIS Validation Study and invite them to participate. Those programs that agreed to participate then worked with the data collector to schedule an observation in one or more classrooms and collect surveys from staff members. Within each program, up to 4 classrooms were randomly selected for observation and child care providers/teachers including aides and other paid staff in those randomly selected classrooms were asked to complete a short Provider Survey and Consent Form. For child care programs with 4 classrooms or fewer, all classrooms were observed and all paid staff were asked to complete the survey and consent form. All data collectors were trained to use the observational tool (CLASS) and were reliable according to CLASS standards as well as reliable with other observers on the data collection (see Measures). On the day of the observation, PSU observers collected the Provider Surveys and consent forms from paid staff in the randomly selected classrooms at the child care program. See additional description of measures and observation procedures below.

As a "thank you" for participating, each program received an Amazon gift card(s). Programs received a $20 Amazon gift card for each classroom that was observed (up to $100 total in gift cards for programs with 4 observed classrooms), and Level 1 programs (those not participating in QRIS) received a $150 gift card for their entire program, regardless of number of classrooms observed. Gift cards were mailed to each program approximately one month following the observation. Observations and survey data collection for phase 1 of the QRIS Validation Study were completed by July 2015.

A total of 790 child care programs were identified for the QRIS Validation Study (455 programs participating in QRIS and 335 programs *not* participating in QRIS referred to as "Level 1" programs). Of these programs, the QRIS Validation Study data collection team was able to contact 599 programs, 428 of which were eligible for the QRIS Validation Study. Of those eligible for the study, 312 participated and were observed by the PSU data collection team. The overall participation rate was 73%.

Participation rates for the QRIS Validation Study for programs already participating in QRIS were high (85%). Participation varied by program type. Certified Centers were more likely to participate (93%) than home-based programs (Certified Family 80% and Registered Family 76%). While the PSU Institutional Review Board prohibited PSU data collection staff from asking child care programs why they declined participation, some programs provided a rationale on their own and their responses were recorded by

the team at PSU. Among programs participating in QRIS, reasons for declining to participate in the Study included being frustrated with the QRIS process, not having enough time/resources, or being too busy.

The participation rate for programs that were not engaged in the QRIS (i.e., Level 1 programs) was lower than that for QRIS participating programs (45%). Larger differences in participation rates across different program types were also seen between Level 1 programs compared to differences in participation among QRIS participating programs. The participation rate for Certified Centers (74%) was much higher than the rates for Certified Family (36%) and Registered Family (32%) child care programs. Even with increased incentives to non-participating programs ($150 Amazon gift card), it was difficult to recruit programs at Level 1. Reasons to decline participation by Level 1 programs included: too much time/effort, not interested in participating in a State-run program, and did not want to bother families. Many other programs were planning to apply for the QRIS within 6 months of the initial recruitment call, which contributed to the high number of Level 1 child care programs that were considered ineligible for phase 1 of the QRIS Validation Study. Finally, many Level 1 programs had closed or were in the process of closing.

Data Management

The OSU data management team received four datasets, which included the five sources of data (the CLASS data set contained information from the observations plus information collected via survey).

- **QRIS data set**: TRI provided data on every program that had submitted information related to QRIS up to May 15, 2015. Not all of the 1,187 programs had submitted portfolios by that date.
- **VS-ORO data set**: TRI provided data on professional development qualifications of staff that they had collected from the Oregon Registry Online (ORO) at the time the portfolio was reviewed by TRI. TRI sent VS-ORO data on 2,605 practitioners from 454 programs.
- **CLASS dataset**: PSU provided CLASS observational data on 314 programs (2 were excluded as they lost their rating due to noncompliance with licensing standards). PSU also provided a file of survey data collected from staff at the time of the observation. These data were not used in Study 1 other than to describe the sample of teachers/providers who were observed.
- **Structural Indicator dataset**: The Hallie Ford Center provided Structural Indicator (SI) data on 4,024 regulated centers and family child care homes that had licenses renewed in 2014. This dataset included measures of six program characteristics related to quality: education, training, wages, benefits, retention, and accreditation. Education and training variables are based on ORO data at the time of licensing renewal.

The OSU data management team created an analysis dataset by merging the three facility-level datasets: QRIS, CLASS, and Structural Indicator. About 60% of the programs in the TRI dataset did not match with PSU dataset, the major reason being that the TRI dataset included large numbers of programs that had not completed the portfolio evaluation process or earned a rating of 2 prior to May 15, 2015. Of the 314 programs that did match, 8 had incomplete QRIS rating data as they had incomplete/unrated portfolios after multiple requests for evidence, and 2 programs had their star-Level revoked due to a compliance issue that happened after rating. As noted in the Procedures section, contact for information for programs that had been identified as Level 1 came from the Hallie Ford Center at OSU. The data management team sent these data to PSU. These programs had not submitted information to TRI so were not in their dataset. The Hallie Ford Center at OSU provided the OSU data management team Structural Indicator data on 4,024 programs, the regulated programs for which they had Structural Indicator data. Ninety-one percent of the 304 programs (277 programs) with overall ratings (level 1-5) had Structural Indicator data. Included in reasons that the 27 (304-277) programs did

not match included that some had changed type of care since the time of their 2014 license renewal and that others had been excluded from the Structural Indicator data due to a change of location. Of the 304 programs in the analysis dataset 58 were Level 1 and 246 had ratings of 2-5.

In order to create an independent measure of personnel qualifications for the Validation Study, OSU used the VS-ORO data collected by TRI to create personnel qualification variables at the person level (captured in a dataset called VS-ORO). Both TRI and PSU sent person-level as well as program-level data. There were problems linking the person-level (practitioners) data due to differences in the unique identifier used in the two databases. To resolve this issue the data management team attempted to link by teacher/provider name. PSU amended their IRB so that OSU could have access to the names of staff. OSU then sent a list of license numbers of the programs to the Early Learning Division (ELD). ELD used the license number to retrieve the names and ORO identification number (ids) of all staff associated with that number from the child care licensing database. OSU matched the names with the PSU (CLASS) and TRI (QRIS) data and attached ORO ids for those that matched. Given the difference in time between portfolio submission and ELD sending staff names and ORO ids, not all individuals were matched with their ORO id.

The OSU data management team cleaned each dataset, converted text variables to numeric, and checked for inconsistencies in the ranges. This step included converting QRIS indicator variables to the names originally created through a cross-walk that was verified with TRI. The data manager combined all the data into one large dataset. She also matched staff in QRIS and PSU databases using the data provided by ELD and included the unique VS-ORO id in the QRIS database when a match was found. At this point the data manager created the variables needed for analyses while continuing to clean and correct the data. Different analysis datasets were created at the program, classroom, and practitioner levels. For example, to create program level variables, practitioner data would be averaged across all practitioners in a given program using the license number. When the 2014 Structural Indicator database became available, the data manager merged those data with the appropriate analysis dataset using the license number. Thus, the Validation Study had three measures of personnel qualifications: QRIS ratings specific to the Personnel Qualifications domain, VS-ORO measures, and Structural indicator education and training variables. Each relied on ORO data but each had unique measures. For the 246 rated programs, all but 14 of these programs had ORO data for at least some, if not all, of their staff; 13 of the 14 were family child care and 1 a center. Thus, we had VS-ORO data on at least some of the staff for 234 programs.

Table 1 displays the relationship among the four datasets. CLASS and Structural Indicator datasets included all or most of the Level 1 programs as well as the level 2-5 programs, whereas QRIS and VS-ORO included only the level 2-5 programs.

Table 1. Effective sample by data source

	CLASS	Structural Indicators	QRIS	VS-ORO
CLASS	304	277	246	234
Structural Indicators	277	277	246	234
QRIS	246	246	246	234
VS-ORO	246	246	246	234

Note. Samples sizes vary by data source. CLASS includes data on 304 programs: 58 Level 1 programs and 246 programs with QRIS ratings 2-5. QRIS contains data on 246 programs with QRIS ratings 2-5 and VS-ORO dataset contains data on 234 of these programs. Structural Indicators includes data on 277 programs: 51 Level 1 programs and 226 programs with QRIS ratings 2-5.

Measures

QRIS Ratings. For the purposes of this study, QRIS ratings ranged from 1 to 5, even though Oregon's QRIS only fully rated programs at the 3-, 4-, and 5-star levels. 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. Level 2 programs had applied for but did not achieve a 3-star rating or higher. Programs that resubmitted their portfolio for a higher rating during the 15-month QRIS Validation Study recruitment period were recruited for, and thus participated in, the Validation Study only one time.

The QRIS ratings included an overall rating (1, 2, 3, 4, or 5), and for the 246 programs that were rated a 2-5 we had ratings for each of the specific standards in the five domains that collectively comprised the overall rating. Additionally, some limited information about the evidence programs submitted to meet specific indicators or aspects of the standards was also available and utilized when appropriate.

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, 2011 and Pianta, La Paro, & Hamre, 2008, respectively). Each tool was used in classrooms/groups where the majority of children (i.e., greater than 66%) were in the tool's age range. For classrooms/groups that consisted of a mix of toddlers and preschoolers, a third tool ("Combined CLASS") was used. This tool was created based on work by Gail Joseph at the University of Washington for the SEEDS project (Joseph, Feldmen, Brennan, Naslund, Phillips, & Petras, 2011) and a cross-walk between the Toddler and PreK CLASS tools. The Combined CLASS tool was used in classrooms where between one-third to two-thirds of the children were from either the toddler or preschool age group.

The Toddler CLASS tool separates adult-child interactions into 2 domains (Emotional and Behavioral Support and Engaged Support for Learning), and is comprised of a total of 8 dimensions. The Emotional and Behavioral Support domain consists of 5 dimensions: (1) positive climate; (2) negative climate; (3) teacher sensitivity; (4) regard for child perspectives; and (5) behavior guidance. The Engaged Support for Learning domain included 3 dimensions: (1) facilitation of learning and development; (2) quality of feedback; and (3) language modeling.

The PreK CLASS tool is comprised of 3 domains (Emotional Support, Classroom Organization, and Instructional Support) with 10 total dimensions. Within the Emotional Support domain, the PreK CLASS tool included the following 4 dimensions: (1) positive climate; (2) negative climate; (3) regard for student perspectives; and (4) teacher sensitivity. The Classroom Organization domain included 3 dimensions: (1) behavior management; (2) productivity; and (3) instructional learning formats. Three dimensions made up the Instructional Support domain: (1) concept development; (2) quality feedback; and (3) language modeling.

Based on a description of the Combined CLASS tool created by Gail Joseph and a cross-walk of the Toddler and PreK CLASS tools, the Combined CLASS tool was broken down into 3 domains (Emotional Support, Classroom Organization, and Instructional Support). The Combined CLASS tool consisted of 11 dimensions of adult-child interactions. These dimensions were the same as those in the PreK CLASS tool, with the exception of Facilitation of Learning and Development, which was added to the Instructional Support domain of the Combined CLASS tool for toddlers only. In addition to this dimension, which was scored only for toddlers in the classroom, 2 dimensions were scored for

preschoolers only (instructional learning formats in the Classroom Organization domain and concept development in the Instructional Support domain). For details on the Combined CLASS tool, see Combined CLASS Behavioral Markers in Appendix D.

Because of the strong parallels between the Toddler domain of Emotional and Behavioral Support with the PreK and Combined domain Emotional Support, and the need for consistency in presentation, the presentation of findings throughout this report uses the term "Emotional Support" to represent the Toddler domain of "Emotional and Behavioral Support." Similarly, we use the term "Instructional Support" to represent not only the PreK and Combined domain of Instructional Support but also the Engaged Support for Learning domain from the Toddler CLASS tool.

Observations using one of the three versions of the CLASS were conducted in up to 4 randomly selected classrooms within each program. Each observation consisted of three observation cycles each lasting 20 minutes. As per guidelines in the CLASS tool manuals, most classroom activities were observed, excluding nap and bathroom time as well as outdoor time for the PreK and Combined CLASS tools. Within each of the 3 versions of the CLASS tool, dimensions were scored on a 7-point scale from 1 (very low) to 7 (very high). Ratings of 1 or 2 are characterized as in the "low range," 3 to 5 in the "mid-range," and 6 to 7 in the "high range" although this study utilizes the 1-7 scores.

The PreK CLASS tool has been found to be a valid tool for assessing adult-child interactions and to have good inter-rater reliability (La Paro, Pianta, & Stuhlman, 2004). Less research has been conducted to assess the reliability and validity of the Toddler CLASS; however, it was developed based on foundational principles for learning and development in young children as well as domains found to be reliable and valid within the PreK CLASS tool (Early et al., 2007; Hamre & Pianta, 2007; Morrison & Connor, 2002; Pianta, La Paro, Payne, Cox, & Bradley, 2002; Rieber, 1998; Rutter & Maughan, 2002).

Training and Reliability**.** All data collectors were trained by Teachstone on the Toddler and PreK CLASS tools and met Teachstone's reliability requirements for CLASS certification (i.e., 80% of codes matching standard codes set by Teachstone and no dimensions with 3 or more ratings consistently scored higher or lower than Teachstone's standard code). Data collectors were also trained on the Combined CLASS tool by the data collection coordinator. Inter-rater reliability within the data collection team was also established at the beginning and middle of the data collection period using the same standards as those set by Teachstone on all three tools. Inter-rater reliability was achieved by pairs of data collectors in the field. On a bi-weekly basis, data collectors met with the data collection coordinator to discuss observations and scoring issues. One year after initial CLASS certification, data collectors were required to re-certify as CLASS observers through Teachstone by completing additional reliability testing.

Calculating Class Scores. To create program-level CLASS scores to examine links with QRIS ratings the scores for each classroom/group observed with the same tool (Toddler, PreK, Combined) were averaged within each program. Additionally, a total average CLASS score for each domain was computed by averaging scores within each program across all of the CLASS instruments with which they were observed. The Total Emotional Support score was composed of Toddler Emotional and Behavioral Support, PreK Emotional Support, and Combined Emotional Support. The Total Instructional Support score was composed of Toddler Engaged Support for Learning, PreK Instructional Support, and

Combined Instructional Support. The Total Organized Classrooms (or "Organizational Support" for consistency with the other two domains) was comprised of the Organized Classrooms scores for the PreK and Combined CLASS tools; Toddler CLASS does not have an equivalent.

Oregon Registry Online (ORO) Data. ORO Registry Online is a statewide database of persons that are employed in child care and education. Through nightly data sharing between the Oregon Center for Career Development at PSU and the Office of Child Care, Early Learning Division, each person who works in a regulated child care facility is linked to the facility in which they are currently employed. The database stores submitted training and education and verifies it for system use, such as Office of Child Care licensing needs and the Department of Human Services (DHS) Enhanced Rate Program. In addition to data on an individual's education and training, ORO contains demographic data on each person.

Structural Indicators of Quality. In September 2001 a team of researchers met to identify indicators that research would predict to be associated with quality (Weber & Wolfe, 2003). The list included: teacher education, teacher training, teacher retention, teacher compensation—wages and benefits, and accreditation. Partners identified data sources and methods for accurately measuring the indicators. Data sources included a) data collected by Child Care Licensing Specialists at the time of licensing renewal visits and managed by the Early Learning Division; b) data stored in the Child Care Regulatory Information System (CCRIS) managed by the Early Learning Division; and c) ORO. Researchers at the Hallie Ford Center for Children and Families at OSU retrieved data from their sources, merged data, and created indicators for each regulated facility (Certified Centers, Certified Family child care, and Registered Family child care). The Hallie Ford Center researcher provided the data to the data management team.

Sample Description

The Validation Study sample included 312 programs. Eight programs had incomplete portfolios and were dropped from the sample, leaving 304 programs for analyses examining links between CLASS scores and QRIS ratings that included Level 1 programs. Some analyses were only possible to conduct with a subsample of programs with QRIS rating data (levels 2-5) because they examined programs' actual scores and data submitted as part of the portfolio process. These analyses utilized the 246 programs that had data from both CLASS observations and QRIS ratings.

The sample represented all three of the child care license designations in Oregon: 65 programs (21%) were designated as Registered Family (RF); 94 programs (30%) were designated as Certified Family (CF); and 153 programs (49%) were designated as Certified Centers. Observed programs served children between the ages of 15 to 60 months (i.e., toddlers and preschoolers). Table 2 indicates the number of programs with at least one class/group by age group of children.

Table 2. Programs with at least one class/group of the following age groups by program type

	Registered Family	Certified Centers	Certified Family
Toddlers	16	87	26
Preschoolers	28	135	38
Toddlers & Preschoolers (Combined)	26	15	42

Programs in the sample ranged in size from those with only a single group/classroom to centers with up to 25 classrooms. As described previously, up to 4 classrooms/groups were observed in each program. Table 3 shows the number of classrooms/groups that were observed in each program by program type. Just over half of the sample (56%) had only one classroom/group observed.

Table 3. Number of classroom/groups per program in sample

# classes/ groups observed	# (%) of programs	Registered Family	Certified Centers	Certified Family
1	175 (56%)	60	33	82
2	50 (16%)	5	36	9
3	36 (12%)		33	3
4	51 (16%)		51	

Staff members (N = 1,084) who were part of the CLASS observations in this study were 96% female. Their positions were as follows: 43% lead/head teacher, 28% assistant teacher, 12% director/owner, 8% assistant/aide, and 9% other. The racial/ethnic background of these staff members was: 79% White, 12% Hispanic, 5% Asian, 3% Black, 2% American Indian, 1% Hawaiian, and 2% other. Eighty-four percent reported English as their primary language; 6% reported Spanish and 5% reported another primary language. Ninety-three percent reported speaking English most often with the children; 2% reported speaking Spanish most often with the children and 5% reported speaking another language most often with the children.

Data Analysis

Analysis was conducted at Oregon State University, with support of the entire Validation Study team. Additionally, in the final phases of this Validation Study 1, the Validation Team partnered with a review team of experts to discuss early findings, consider possible interpretations of the data, and to identify additional analyses to further examine the data. This team, referred to as the QRIS Validation Study "mini review team" represented the QRIS Implementation Team, the QRIS Process Evaluation Team, the Early Learning Division, the QRIS Technical Assistance Specialists, and Oregon Center for Career Development staff. Specific analytic approaches are described in each relevant section of the Results.

Results

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

Summary of QRIS Ratings

Of the 304 programs participating in the QRIS Validation Study, 19% were Level 1 programs and 81% were rated by the QRIS at levels 2 through 5. Also, approximately one-half (49%) of the 304 programs were Centers and 51% were family child care (21% Registered and 30% Certified). Of the 246 programs with QRIS ratings, 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%) (Table 4). These percentages varied substantially by type of care. For example, nearly one-half Registered Family providers were rated at star-Level 3, with only 6 programs (13%) rated at star-levels 4 or 5. Centers and Certified Family providers had higher proportions of programs at star-levels 4 -5 (36% for Centers; 40% for Certified Family). Almost one-fifth of the overall sample (N = 58) was comprised of the Level 1 programs that were recruited as a likely "low quality" comparison group of programs not participating in the QRIS.

Table 4. Programs by QRIS rating and program type

QRIS Rating	Total Programs	Registered Family	Certified Center	Certified Family
Total	**304 (100%)**	**63 (21%)**	**149 (49%)**	**92 (30%)**
Level 1	**58 (19%)**	**18 (31%)**	**29 (50%)**	**11 (19%)**
QRIS Rating 2-5	**246 (81%)**	**45 (18%)**	**120 (49%)**	**81 (33%)**
Among QRIS Rated Programs: # (%) of programs rated 2-5				
2	91 (37%)	17 (38%)	56 (47%)	18 (22%)
3	74 (30%)	22 (49%)	21 (18%)	31 (38%)
4	23 (9%)	4 (9%)	8 (7%)	11 (14%)
5	58 (24%)	2 (4%)	35 (29%)	21 (26%)

*Level 1 was identified by the data analysis team through the Structural Indicator data, and were not rated through the QRIS.

Note. 8 of the original 312 programs in the Validation Study sample did not have a rating available, so they were excluded from the analysis.

Table 5 shows the percentages of the 246 QRIS-rated programs by their rating for each standard within the QRIS, organized by domain (average ratings are available in Appendix E). Two overall patterns emerge from this table. First, the percentage of programs with a star-3, 4, or 5 on individual standards is almost always higher than it is for the percentage of programs rated star-3, 4, or 5 overall on the QRIS. In other words, programs are doing better on some standards than is reflected in their overall QRIS rating. Second, some standards are harder for programs overall, as evidenced by large percentages of programs at Level 2 and/or small percentages of programs at levels 3-5.

Tables depicting these percentages by program type are available in Appendix E and show a similar pattern to the one in Table 5, in which higher proportions of Registered Family programs were rated 2 and 3, and higher proportions of Centers and Certified Family providers were rated star-4 and 5.

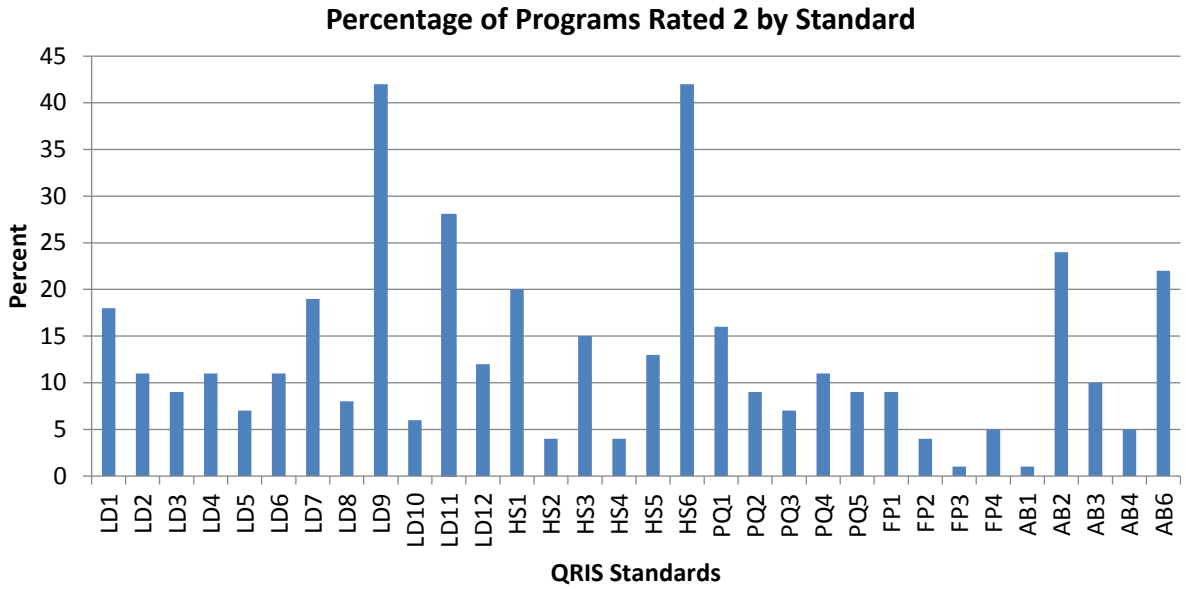
Table 5. Frequencies of QRIS Ratings (all programs)

Domain & Standards	Total	Missing (N/A)	# (%) Rated at each star-level			
			2	3	4	5
Learning & Development 1	246	0	44 (18%)	87 (35%)	34 (14%)	81 (33%)
Learning & Development 2	246	0	28 (11%)	85 (35%)	26 (11%)	107 (43%)
Learning & Development 3	246	0	21 (8%)	106 (43%)	36 (15%)	83 (34%)
Learning & Development 4	246	0	28 (11%)	103 (42%)	29 (12%)	86 (35%)
Learning & Development 5	246	0	16 (7%)	91(37%)	30 (12%)	109 (44%)
Learning & Development 6	246	0	26 (11%)	104 (42%)	31(13%)	85 (34%)
Learning & Development 7	246	0	46 (19%)	80 (32%)	31 (13%)	89 (36%)
Learning & Development 8	246	0	22 (9%)	98 (40%)	32 (13%)	94 (38%)
Learning & Development 9	246	0	104 (42%)	68 (28%)	12 (5%)	62 (25%)
Learning & Development 10	246	0	15 (6%)	108 (44%)	34 (14%)	89 (36%)
Learning & Development 11	246	0	69 (28%)	78 (32%)	58 (24%)	41 (17%)
Learning & Development 12	246	0	29 (12%)	107 (44%)	43 (17%)	67 (27%)
Health & Safety 1	246	0	50 (20%)	80 (33%)	32 (13%)	84 (34%)
Health & Safety 2	246	0	10 (4%)	117 (48%)	25 (10%)	94 (38%)
Health & Safety 3	246	0	38 (15%)	100 (41%)	27 (11%)	81 (33%)
Health & Safety 4	246	0	10 (4%)	111 (45%)	27 (11%)	98 (40%)
Health & Safety 5	246	0	32 (13%)	97 (39%)	26 (11%)	91 (37%)
Health & Safety 6	246	0	103(42%)	64 (26%)	26 (11%)	53 (21%)
Personnel Qualifications 1	246	0	40 (16%)	85 (35%)	33 (13%)	88 (36%)
Personnel Qualifications 2	246	86(35%)	23 (9%)	46 (19%)	20 (8%)	71 (29%)
Personnel Qualifications 3	246	93(38%)	18 (7%)	46 (19%)	22 (9%)	67(27%)
Personnel Qualifications 4	246	0	27 (11%)	82 (33%)	36 (15%)	101(41.06)
Personnel Qualifications 5	246	0	21 (9%)	114 (46%)	23 (9%)	88 (36%)
Family Partnerships 1	246	0	22 (9%)	107 (43%)	49 (20%)	68 (28%)
Family Partnerships 2	246	0	13 (5%)	126 (51%)	29 (12%)	78 (32%)
Family Partnerships 3	246	0	2 (1%)	102 (42%)	45 (18%)	97 (39%)
Family Partnerships 4	246	0	12 (9%)	110 (45%)	29 (12%)	95 (39%)
Admin & Business Practice 1	246	2 (1%)	9 (4%)	115 (47%)	22 (9%)	98 (40%)
Admin & Business Practice 2	246	59 (24%)	16 (6%)	63 (26%)	20 (8%)	88 (36%)
Admin & Business Practice 3	246	60 (24%)	24 (10%)	67 (27%)	26 (11%)	69 (28%)
Admin & Business Practice 4	246	61 (25%)	11 (4%)	62 (25%)	22 (9%)	90 (37%)
Admin & Business Practice 5	246	1 (<1%)	52 (21%)	86 (35%)	28 (11%)	79 (32%)
Admin & Business Practice 6	246	58 (24%)	5 (2%)	64 (26%)	27 (11%)	92 (37%)

Note. Minimum score for all standards is 2 and maximum is 5.

Figure 1 illustrates this pattern graphically. Standards with more than one-quarter of programs scoring a 2 are Learning and Development (LD) 9 and 11, and Health and Safety (HS) 6, although several other standards also have relatively high percentages of programs scoring a 2. Reviewing these patterns by type of care (see Appendix E) reveals additional standards for which more than a quarter of Registered Family providers scored a 2: LD1, LD7, HS1, HS3, HS6, Personnel Qualifications (PQ) 1, and Administration and Business Practices (AB) 5.

Figure 1. Percentage of programs rated 2 by standards in each domain



Abbreviations for domains are as follows: Learning and Development (LD), Health and Safety (HS), Personnel Qualifications (PQ), Family Partnerships (FP), and Administration and Business Practice (AB).

Summary of CLASS Scores

The average CLASS scores across all groups/classes observed within each program are presented in Table 6, for each CLASS instrument (PreK, Toddler, Combined), including each of the CLASS domains and the total score. The means of these program-level average CLASS scores are around 5 for Emotional Support, 4.5 for Organizational Support, and 2.5 for Instructional Support, with an overall total average around 4 (on a scale from 1 to 7). These scores are similar to those documented in other studies using the CLASS (Hatfield, Burchinal, Pianta, & Sideris, 2016; Burchinal, Vandergrift, Pianta, & Mashburn, 2010).

Despite the relatively high average scores for Emotional Support and low average scores for Instructional Support, the minimum and maximum scores, coupled with the Standard Deviation indicate substantial variability in programs scores. For example, programs scored as low as 1.0 and as high as 6.0 on Instructional Support. They also scored as low as 2.85 and as high as 6.83 on Emotional Support.

Table 6. Descriptive statistics for CLASS scores (all programs)

CLASS Scores Average	N	Minimum	Maximum	Median	Mean	SD
Total Average Across All Instruments						
Emotional Support[1]	312	2.85	6.83	5.42	5.37	0.73
Organized Classrooms[2]	267	2.11	6.78	4.67	4.64	0.83
Instructional Support	312	1.00	6.00	2.56	2.66	0.82
Total	312	2.29	6.27	4.24	4.28	0.69
PreK CLASS						
Emotional Support	201	3.00	7.00	5.56	5.43	0.69
Organized Classrooms[2]	201	2.00	7.00	4.78	4.71	0.85
Instructional Support	201	1.00	5.00	2.33	2.51	0.79
Total	201	2.00	6.00	4.21	4.22	0.67
Toddler CLASS						
Emotional & Behavioral Support	129	2.85	6.88	5.21	5.12	0.92
Instructional Support	129	1.00	7.00	2.78	2.91	0.90
Total	129	2.00	7.00	4.39	4.38	0.81
Combined CLASS for mixed age-groups						
Emotional Support	83	3.00	7.00	5.50	5.37	0.74
Organized Classrooms[2]	83	3.00	7.00	4.50	4.53	0.85
Instructional Support	83	1.00	5.00	2.50	2.52	0.81
Total	83	2.00	6.00	4.13	4.14	0.70

[1]Emotional Support for Toddler Measure Includes Behavioral Guidance

[2]Toddler Measure does not Include Organization Support. Behavioral Guidance included in Emotional Support

These patterns of CLASS scores were similar across all three types of care: Centers, Certified Family, and Registered Family even though most of the Registered Family programs were rated 1-3 (Appendix E).

In sum, of the 246 programs with QRIS ratings (2 through 5), over one-third (37%) were star-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%).

Overall program-level CLASS scores were in the upper end of the "mid" range for Emotional Support and Organizational Support, and at the upper end of the "low" range for Instructional Support. 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.

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

The five domains of the QRIS were highly correlated (Table 7). These correlations were so large (range from.82 to.94) that they were close to 1.0, which is the maximum possible value for a correlation, indicating that the two variables measure exactly the same underlying construct (e.g. they were essentially measuring the same thing).

Table 7. Correlation among QRIS domains

	Domains	Learning Development	Personnel Qualifications	Health & Safety	Family Partnerships	Administration & Business Practice
Total	Learning Development	1.00				
	Personnel Qualifications	.85***	1.00			
	Health & Safety	.94***	.84***	1.00		
	Family Partnerships	.92***	.82***	.89***	1.00	
	Administration & Business Practice	.91***	.82***	.89***	.86***	1.00

*** Correlation is significant at the p <.001 level (2-tailed)

Similarly, the correlations among standards within each domain are sizeable and all are statistically significant. They range from .54 to .87 for the standards within the Learning & Development domain, from .43 to .78 for Health and Safety, from .60 to .71 for Personnel Qualifications, .77 to .87 for Family Partnerships, and .65 to .83 for Administration and Business Practices (see Appendix E). The smaller correlations represented in this summary (e.g. .43, .54, etc.) involve LD9 (screening) and HS6 (screen time). Appendix E also summarizes the alpha coefficients, which represent the internal consistency within each of the five QRIS domains. They all exceed .90, indicating very high consistency among programs' scores on the various standards within the QRIS.

These large correlations are likely due to the five domains having been packaged within an overall portfolio that programs submit to demonstrate they have reached specific standards at a consistent level (e.g. all standards at a 3-star Level). If the domains had been measured separately from one another, and/or in a way that captured the full variability of programs' practices on each standard rather than essentially truncating variability at the level for which programs applied, the correlations would likely be substantially smaller. Although programs were encouraged to submit evidence for standards at higher levels than the one for which they were applying, few did.

Evidence from prior studies in the field consistently point to much smaller correlations among aspects of early learning programs that are measured by Oregon's QRIS standards and domains. For example, in an analysis of six large existing data sets researcher documented correlations among measures of staff education, training, group size, ratio, curricular practices, family involvement, adult-child interactions ranging from r = .15 to r = .55 (Burchinal et al., 2016).

In sum, the QRIS domains and standards correlations appear artificially high; likely a result of the portfolio/block structure of Oregon's QRIS. These high correlations among the various parts of the QRIS 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.

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

This study took three complementary approaches to examine the primary research question about how well QRIS ratings differentiated observed quality of adult-child interactions, using the CLASS. First, we conducted Pearson's correlations to estimate the size and significance of the associations between QRIS ratings and CLASS scores. It is important to note that correlations assume a linear relationship, such that each increase in a QRIS rating is associated with the same amount of increase in CLASS scores. Next, we conducted Analysis of Variance (ANOVA) tests to detect any differences in CLASS scores based on QRIS ratings. The advantage of the ANOVA is that it allows for detection of non-linear associations (e.g. do programs at QRIS star-level 3-5 score higher on the CLASS than programs at levels 1-2). Finally, we followed up on the results from the correlations and ANOVAs to better understand them, using cross-tabs. With cross-tabs we were able to document the actual number of programs that had high CLASS scores but low QRIS ratings, or low CLASS scores but high QRIS ratings, etc. By triangulating evidence across these three approaches, we gained confidence in the conclusions we drew from the data.

Correlations among CLASS scores on QRIS ratings

Table 8 presents the correlations among programs' QRIS ratings and CLASS scores. CLASS scores are calculated at the program-level. The Overall CLASS scores represent the average score across all groups/classes observed in this study; the PreK scores represent the average score across all groups/classes within the program observed with the PreK version of the CLASS, etc. The sample size for each correlation varied analysis by analysis, depending on the number of programs with one or more groups/classes observed with each instrument (Toddler-CLASS, PreK-CLASS, Combined-CLASS).

Overall, CLASS scores showed small positive correlations with QRIS ratings for the Organizational (r =.19, p < .05) and Instructional (r =.20, p < .05) domains. Correlations were slightly larger (in what is considered the "moderate" range) for the Instructional domain on the PreK CLASS (r =.30, p < .05) and the Organizational domain of the Combined CLASS (r =.30, p < .05).

Overall, CLASS scores showed small positive correlations with QRIS ratings.

Table 8. Correlations among QRIS ratings and each CLASS instrument and domain.

CLASS Instrument	CLASS Domain	Correlation with QRIS Ratings	Number of programs Contributing to this Correlation
Overall	Emotional Support	.10	304
	Organized Classrooms	**.19***	259
	Instructional Support	**.20***	304
	Total	**.16***	304
PreK	Emotional Support	**.17***	195
	Organized Classrooms	.14	195
	Instructional Support	**.30****	195
	Total	**.23****	195
Toddler	Emotional & Behavioral Support	.01	126
	Instructional Support	.09	126
	Total	.03	126
Combined	Emotional Support	**.24***	80
	Organized Classrooms	**.30****	80
	Instructional Support	**.23***	80
	Total	**.30****	80

+ Correlation is nearing significance at the p < .10 level (2-tailed)

* Correlation is significant at the p < .05 level (2-tailed)

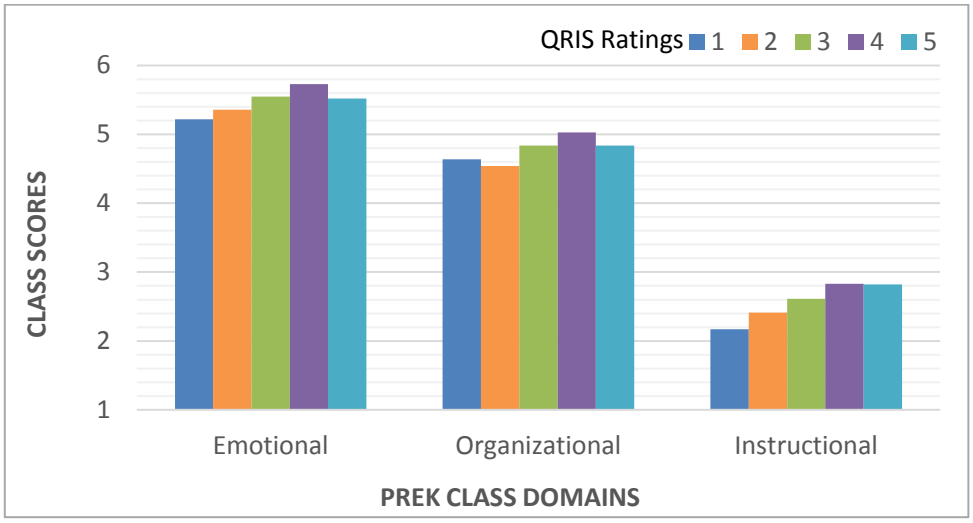
** Correlation is significant at the p < .01 level (2-tailed)

Analysis of Variance: Differences in CLASS scores by QRIS Rating

This report focuses on examining potential differences in CLASS scores between programs rated 1 or2 (combined) versus those rated 3 through 5 (combined) on the QRIS. The rationale for this approach is two-fold. First, efforts related to Oregon's QRIS tend to emphasize achievement of a star rating at level 3, 4, or 5. Although 4- and 5-star ratings are intended to represent higher quality within Oregon's QRIS than a rating of a 3, there is often more of a focus on star ratings of 3-5 collectively, compared to not applying for and/or achieving a rating. Second, within the Validation Study the sample size for certain groups (e.g. programs rated a 4) are too small to adequately compare each QRIS level to each other QRIS level (see Method section).

Preliminary analyses. However, as a preliminary step this study did explore CLASS scores by individual QRIS ratings to help inform the primary analyses, described below. These analyses found that although there were a few instances of a 'stair-step' type pattern of increases in CLASS scores by QRIS rating of 1 vs 2 vs 3 vs 4 vs 5 (see Instructional Support in Figure 2 for a partial example), more often CLASS scores did not increase consistently with each increase in QRIS rating. Many times the CLASS scores for programs rated a 5 were either virtually equivalent to, or lower than scores for programs rated 3 or 4 (see Emotional and Organizational domains of the PreK CLASS in Figure 2). Figure 2 is provided as one illustration of several such exploratory analyses. In general, results indicated no differences between programs rated 1 vs 2, and no differences between programs rated 3 vs 4 or 5, or between programs rated 5 vs those rated 3 or 4.

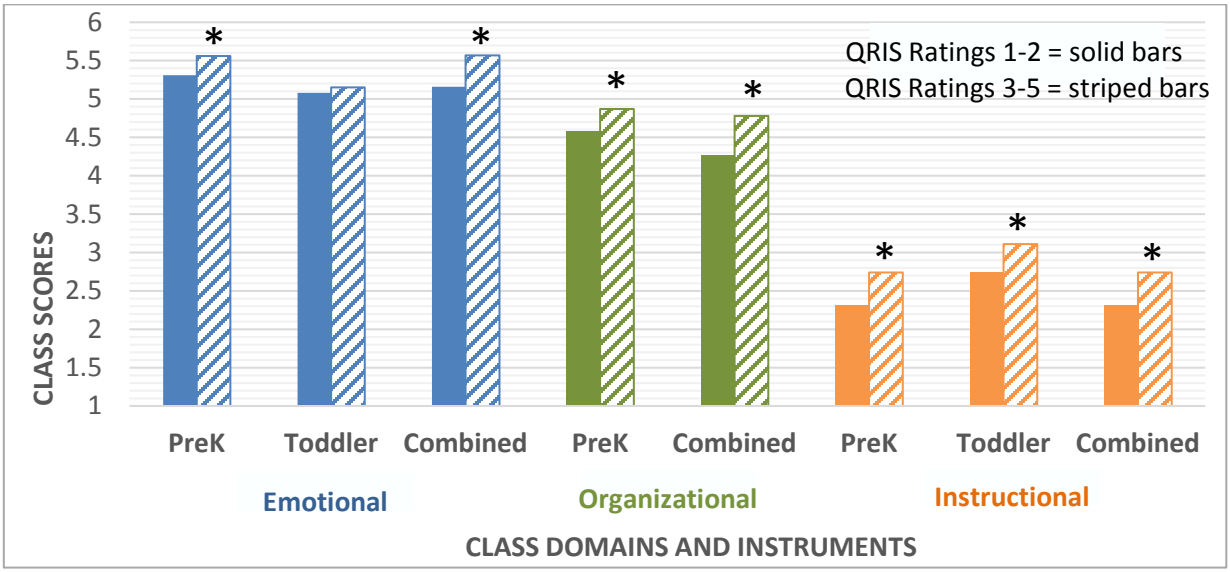
Figure 2. Exploration of PreK CLASS scores by QRIS Ratings 1, 2, 3, 4, and 5 (includes all program types)



* PreK CLASS scores are presented by each QRIS rating (1, 2, 3, 4, and 5) to show patterns, not to illustrate significant differences.

Hypothesis testing. Overall, programs that achieved a star rating at level 3, 4, or 5 showed significantly higher CLASS scores than those rated 1 or 2 (Figure 3).

Figure 3. CLASS scores by QRIS rating (includes all program types)



* Differences between programs rated 1-2 and 3-5 are statistically significant, $p < .05$.

The results from the significance testing for the data presented in Figure 3, as well as those broken down by type of care, are presented in Table 9. Table 9 summarizes the results from several analyses into one table by presenting the F values, which represent the amount of difference in CLASS scores between programs rated 1-2 and those rated 3-5. When combining across program types, those programs rated 3, 4, or 5 score higher on all three domains of both the PreK and Combined CLASS tools than programs rated 1 or 2. Differences in CLASS scores were the largest and most consistently significant (across domains and types of programs) on the PreK CLASS tool. The only difference detected with the Toddler CLASS was for the Instructional domain in Centers. Effect sizes (Cohen's d) for these differences in the overall CLASS scores across all type of programs by QRIS rating were small to medium:.26,.42, and.44 for the Emotional, Organizational, and Instructional domains, respectively (see Appendix E).

Programs that achieved a 3-star rating or higher on the QRIS showed higher-quality adult-child interactions than those rated 1 or 2.

These differences were smaller and less consistent when only comparing programs rated 2 versus 3-star or higher.

Looking specifically at Certified Centers, the only differences detected in CLASS scores by QRIS rating were for the Instructional domain (on both the PreK and the Toddler CLASS). For Certified Family providers, the Organizational domain of the CLASS was the only domain with statistically significant differences by QRIS rating, although the differences on the Instructional domain were close to statistically significant for both the PreK and Combined CLASS tools. There were also differences on the Emotional domain of the PreK CLASS that were nearly significant for Certified Family providers. For Registered Family providers there were statistically significant differences in the Instructional and Organizational domains.

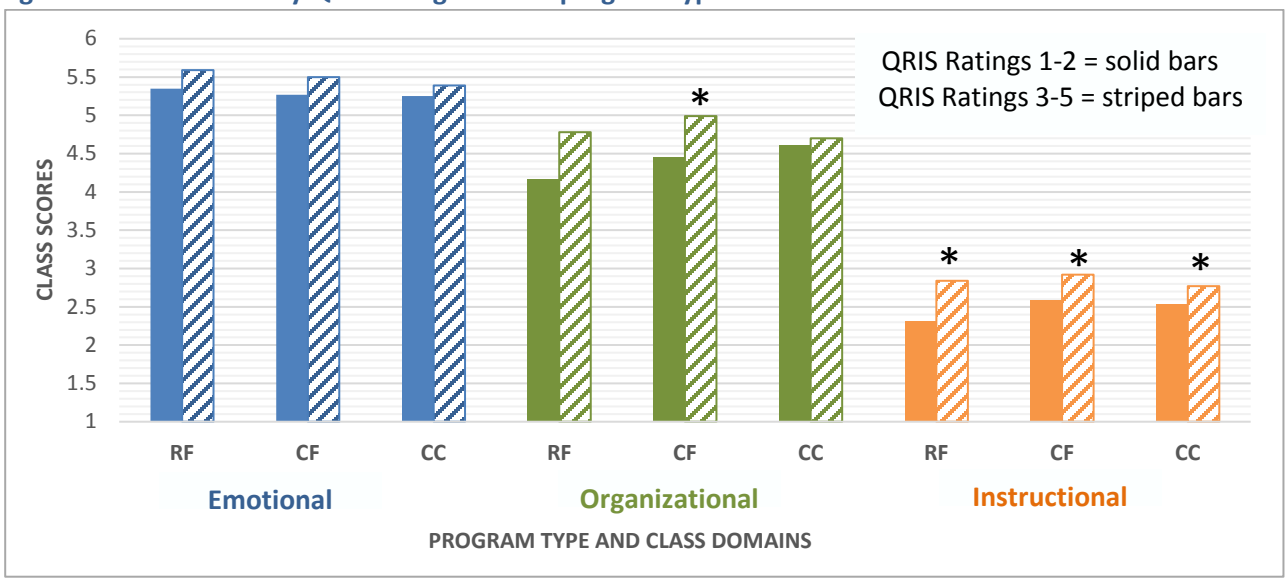
Table 9. Differences in CLASS scores between programs with QRIS ratings of 1 or 2 versus 3-5.

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS Instrument	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Overall												
F	**5.22***	**14.66****	**11.54****	1.40	**3.67+**	0.52	1.75	**3.45+**	**7.26****	1.58	**5.23***	**6.65***
N	304	304	259	149	149	136	92	92	74	63	63	49
PreK												
F	**6.85***	**14.93****	**5.96***	2.46	**5.07***	0.10	**3.75+**	**3.21+**	**4.79***	0.05	**4.15+**	**3.80+**
N	195	195	195	131	131	131	37	37	37	27	27	27
Toddler												
F	0.19	**5.31***	n/a	0.96	**5.26***	n/a	0.00	0.00	n/a	**^**	**^**	**n/a**
N	126	126		84	84		26	26				
Combined												
F	**6.23***	**5.93***	**8.17****	**^**	**^**	**^**	1.04	**2.88+**	**3.69+**	**3.48+**	2.53	2.29
N	80	80	80				41	41	41	25	25	25

Note. Estimates in the table are the F values from the ANOVA tests comparing 1 & 2 level programs vs 3-5 level programs for each CLASS domain.
CLASS domains are Emotional Support (ES), Instructional Support (IS), and Organized Classrooms (OC)
+ Nearing significance at the $p < .10$ level (2-tailed)
* Statistically significant at the $p < .05$ level (2-tailed)
** Statistically significant at the $p < .01$ level (2-tailed)
^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

The differences in CLASS scores by programs rated 1-2 vs 3-5 are similar across the three types of programs. As shown in Figure 3, when looking at the striped bars representing CLASS scores for programs rated 3-5, the scores are very similar across the three types of programs: Registered Family (RF), Certified Family (CF), and Certified Centers (CC). Significance of the differences between programs rated 1-2 vs 3-5 are also very similar across program types. The differences in Instructional Support were statistically significant for each of the three program types. In contrast, the differences in Emotional Support were not statistically significant for any specific type of program, even though there were significant differences in Emotional Support when all types of programs are considered together (Figure 4). The most notable difference by type of program is that Certified Family programs rated 3-5 showed significantly higher CLASS scores on the Organizational domain; this was not the case for Registered Family or Certified Centers.

Figure 4. 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.

Additional analyses were conducted without the Level 1 programs to explore how many of these associations held when comparing the programs rated a 2 with those rated 3 or higher (see Appendix E). Overall, findings from these analyses revealed fewer significant differences in CLASS scores based on QRIS ratings than the analyses that included the Level 1 programs. When grouping all three types of programs together, programs rated 3-star or higher on the QRIS showed significantly higher CLASS scores on the Instructional (all three CLASS tools) and Organizational (PreK and Combined tools) domains. The sizes of the differences were also smaller in the analyses that did not include the Level 1 programs. Additionally, few differences were statistically significant, when looking specifically at each type of program.

CLASS scores for programs rated 3-star or higher on the QRIS are similar across the three types of programs.

Differences in CLASS scores by QRIS ratings are also similar across the three types of programs.

In sum, findings from ANOVAs indicate that programs rated 3, 4, or 5 on the QRIS tended to show slightly higher scores on the CLASS than programs rated 1 or 2 on the QRIS. This pattern of finding was similar across the three program types. There are several possible explanations for a lack of stair-step type pattern of increases in CLASS scores by QRIS rating of 1 vs 2 vs 3 vs 4 vs 5. One possible reason is limited statistical power to detect differences even if they do exist, due to relatively small sample sizes. That said, it is clear that the magnitude of the differences between each individual QRIS rating level are quite small. There were few programs rated 4-star, raising the question of whether this level is meaningful for understanding real differences in program quality. Further, few Registered Family programs achieved 4 or 5- star ratings; among Registered Family programs there were only 6 programs rated either 4- or 5-stars. It is also possible that the QRIS requirements to achieve a 4 and/or 5-star rating are not sufficiently different from those for a 3-star rating to reflect detectible differences in CLASS scores. As with any study, error in either QRIS ratings or CLASS observation scores could also be a contributing factor.

An additional consideration in understanding the validity of QRIS ratings is that each program receives one overall QRIS rating, whereas the quality of care provided to children, reflected by the CLASS scores, varies from one class/group of children to the next, within a program. This is particularly relevant to Centers; only 17 Family providers had more than one group of children for the Validation Study to observe. Results from analysis of multilevel models revealed that only approximately 28% (Organizational) to 43% (Instructional) of the variance in PreK CLASS scores is accounted for by differences between programs. Even when accounting for error, that means that there is a sizeable amount of variation in CLASS scores across classrooms within the same program. This introduces challenges for QRIS ratings of overall programs to reflect the quality of adult-child interactions within classrooms/groups.

> The quality of adult-child interactions varies by classroom/group within programs. This presents challenges for QRIS ratings, which are intended to represent the quality of the overall program.

Cross-tab descriptions of correspondence between QRIS ratings and CLASS scores.

To shed light on why associations between QRIS ratings and CLASS scores were not more substantial, cross-tabs of QRIS ratings and CLASS scores were examined. This follow-up analysis focused on programs that had submitted QRIS portfolios (rated 2, 3, 4, or 5). Cross-tabs are a descriptive tool to summarize the correspondence between QRIS ratings and CLASS scores. They do not test statistical significance; data have to be organized in categories prior to running cross-tabs. QRIS ratings were already in categories of 2, 3, 4, and 5. CLASS scores (which are measured on a continuum, with decimal points), were categorized into "high," "medium," and "low" for the purposes of this analysis. It is important to note that the categories used in this analysis are based on the distribution of CLASS scores within the dataset, not by the categorization of lower, mid, and upper-ranges specified by the developers of the CLASS (La Paro et al., 2011; Pianta et al., 2008). This approach was required because the majority of programs in this study scored in the upper range on Emotional Support and in the lower range on Instructional Support (as defined by the CLASS). To create categories of "high," "medium," and "low" that were meaningful for the current study we used the following cut offs (Table 10).

Table 10. CLASS domains cut-offs (low/medium/high)

	Cut offs for the categories created for this study		
CLASS Domains	**Low**	**Medium**	**High**
Emotional Support	Less than 4.00	5.00 – 5.99	6.00 and higher
Organized Classrooms	Less than 4.00	4.00 – 4.99	5.00 and higher
Instructional Support	Less than 2.00	2.00 – 2.99	3.00 and higher

Table 11 summarizes the correspondence between QRIS ratings and CLASS scores by the number and percentage of programs receiving a QRIS rating of 2, 3, 4, or 5 who had a CLASS score of "low," "medium," or "high." The numbers and percentages highlighted in bold are those that represent a lack of correspondence between QRIS ratings and CLASS scores (e.g. CLASS score is low but program has a QRIS rating of 3 or higher; CLASS score is high but program has a QRIS rating of a 2).

Overall, findings illustrate a mix of good and poor correspondence between QRIS ratings and CLASS scores. For example, consider the Emotional Support (ES) domain. Of the 91 programs with a QRIS rating of 2, 30% had a "low" ES CLASS score, indicating good correspondence between ES CLASS scores and QRS ratings for these providers. Another 48% had a "medium" ES CLASS score, and 22% had a "high" ES CLASS score, which

would indicate poor correspondence between QRIS ratings and CLASS scores for these programs, on this ES domain of the CLASS.

Following the column of QRIS ratings of 2 down through Table 11 shows that between 22% (ES) and 34% (OC) of the programs with a QRIS rating of a 2 actually demonstrated "high" CLASS scores, relative to other programs in the QRIS Validation Study sample. This suggests that at least 20% of the programs with QRIS ratings of a 2 have higher quality of adult-child interactions than are reflected in their QRIS ratings. Additional data tables for each of the three CLASS instrument types show similar patterns for the PreK, Toddler, and Combined versions of the CLASS; the Toddler CLASS showed somewhat higher percentages of the programs rated 2 on the QRIS exhibiting "high" CLASS scores (see Appendix E).

Looking at the programs with high QRIS ratings (4-5), fewer of them have "low" CLASS scores, especially for the Organizational and Instructional domains. Recall that programs scored strongly on the Emotional domain overall; thus, most programs with "low" ES within this sample actually score in the mid-to-upper ranges on the ES domain of the CLASS overall.

Table 11. For programs with each QRIS rating what number (%) had low, medium, and high CLASS scores? (averaged across the PreK, Toddler, and Combined CLASS)

	QRIS Ratings			
CLASS	**2**	**3**	**4**	**5**
Emotional Support (ES)				
Low	27(30%)	**14(19%)**	**6(26.1%)**	**13(22.4%)**
Medium	44(48%)	41(55%)	10(43.5%)	35(60.3%)
High	**20(22%)**	19(26%)	7(30.4%)	10(17.2%)
Total	91(100%)	74(100%)	23(100%)	58(100%)
Organized Classrooms (OC)				
Low	17(19%)	**9(12%)**	0(0%)	**10(17.3%)**
Medium	43(47%)	30(41%)	9(39%)	14(24.1%)
High	**31(34%)**	35(47%)	14(61%)	34(59.6%)
Total	91(100%)	74(100%)	23(100%)	58(100%)
Instructional Support (IS)				
Low	17(19%)	**8(11%)**	**3(13%)**	**6(10.3%)**
Medium	53(58%)	36(49%)	10(43.5%)	28(48.3%)
High	**21(23%)**	30(40%)	10(43.5%)	24(41.4%)
Total	91(100%)	74(100%)	23(100%)	58(100%)

Notes. The cut-off points used to create high, medium, and low CLASS scores were based on the distribution of the dataset values rather than by the categorization of high, medium, and low created by the creators of the CLASS. The numbers and percentages highlighted in bold are those that represent a lack of correspondence between QRIS ratings and CLASS scores (e.g. CLASS score is low but program has a QRIS rating of 3 or higher; CLASS score is high but program has a QRIS rating of a 2).

20% to 30% of the programs rated a 2 on Oregon's QRIS had among the highest CLASS scores in the study (6 or higher on Emotional, 5 or higher on Organizational, 3 or higher on Instructional).

In sum, Table 11 shows more programs with high CLASS scores rated low on the QRIS than the reverse (programs with low CLASS scores rated high on the QRIS). Twenty to thirty percent of the programs rated a 2 on Oregon's QRIS had among the highest CLASS scores in the study (6 or higher on Emotional Support, 5 or higher on Organized Classrooms, 3 or higher on Instructional Support).

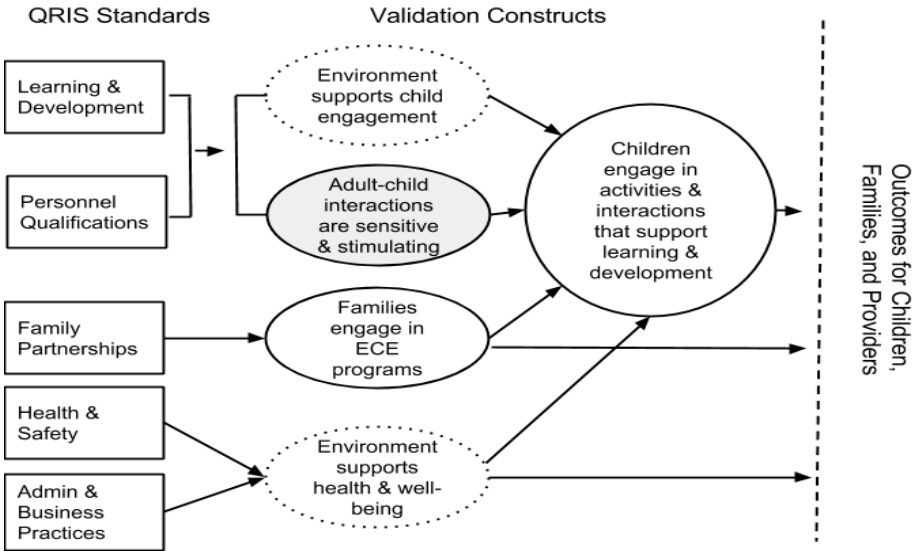
Programs that achieved a 3-, 4-, or 5- star rating had significantly higher quality adult-child interactions, as measured by the CLASS, than those rated 1 or 2. However, these differences varied somewhat by type of program, and the age group of children or CLASS tool used. Additionally, Oregon's QRIS appears to keep low observed quality programs from getting a high rating but keeps over 20% to 30% of programs with high observed quality from getting a high QRIS rating.

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

The current study also completed exploratory analyses to begin to understand how specific standards within the QRIS might relate to programs' CLASS scores. This part of the work is highly exploratory and should be interpreted with caution; the structure of the QRIS leads standards to be highly related to one another (see Results Section 2). Thus, it is completely possible that associations between a given QRIS standard and CLASS scores could actually be due to something else (e.g. other standard(s) within the QRIS). Nonetheless, given the pressing need to provide some information to consider along with other sources beyond the Validation Study for the revision of Oregon's QRIS, we proceeded with this analysis.

This part of the Validation Study narrows in on the two domains (Learning & Development and Personnel Qualifications) of the QRIS that are the most theoretically aligned with the outcome measured in this study: adult-child interactions. This alignment is illustrated in the conceptual map of Oregon QRIS Standards to Validation Study Constructs (Figure 5) created in consultation with the Oregon QRIS Implementation Team in 2014.

Figure 5. Oregon Map of QRIS Standards to Validation Study Constructs



Rectangles: QRIS domains, included in both Validation Study 1 and 2.
Shaded oval: outcome included in Validation Study 1 (CLASS ratings).
Solid line ovals and circle: outcomes included in Validation Study 2 (not part of this report).
Dotted ovals: possible outcomes, not included in the Validation Study.

We used three complementary analytic approaches to examine how programs' ratings on specific QRIS standards relate to their CLASS scores: 1) cross-tabs of the correspondence between the QRIS ratings on specific standards and CLASS scores; 2) Pearson's correlations to examine associations between QRIS ratings on specific standards and CLASS scores; and 3) Analysis of Variance (ANOVA) tests to detect any differences in CLASS scores based on QRIS ratings on specific standards. Correlations assume a linear relationship between variables, such that each increase in a QRIS rating (e.g. from a 2 to a 3 and a 3 to a 4 etc.) is associated with an equal amount of increase in CLASS scores, in a stair-step type fashion. The Analysis of Variance (ANOVA) tests whether there are any differences in CLASS scores across programs with different QRIS ratings. A significant ANOVA test means that there are differences between programs with different ratings, but does not identify which ratings are different from the others (follow

up tests have been conducted to examine this where appropriate). Although work that focuses on specific standards is exploratory, by triangulating evidence across these three approaches we gain confidence in the conclusions we draw from the data.

Only programs rated 2 through 5 on the QRIS are included in the analyses because Level 1 programs do not have QRIS ratings. The results from these analyses of the standards that comprise the Learning and Development and Personnel Qualifications domains are extensive. Thus, we present a summary of the results in this section. Findings for each of these 16 standards is available in Appendix E. Note, the Validation Study does not report associations between programs PQ5 (ethics) ratings and their CLASS scores because the PQ5 standard, and the evidence programs must submit to achieve it, differs substantially from the other PQ standards, and does not have a strong theoretical link with the quality of adult-child interactions.

Summary of Findings: Associations among QRIS standards and CLASS scores

Overall, findings from this exploratory work were fairly similar to the findings for the overall QRIS presented in Results Section 3. Findings pointed to some small, significant links between specific QRIS standards and CLASS scores. Given the high correlations among the QRIS standards, it is not possible for these analyses to identify "the few and powerful" standards. No standards had strong or "powerful" associations with CLASS scores. Yet, analyses did reveal a number of concerns regarding specific standards that may be important to consider, alongside other sources of information, in efforts to strengthen Oregon's QRIS. More detail is available in Appendix E.

Table 12 identifies standards that are of concern and/or which may warrant further consideration for revisions of the QRIS system. This table summarizes the standards that have either no significant links with the CLASS ("0") or in which higher quality programs receive ratings of a 2 on the QRIS standard ("x"). Specifically, the table uses a "0" to indicate instances in which there were *no* significant links between the standard and the CLASS domain on any of the three CLASS tools (PreK, Toddler, Combined) from correlations and ANOVAs. An "x" denotes instances in which more than 20% of programs (1 out of every 5) with a rating of 2 on the QRIS standard scored high on this CLASS domain. These two indicators are conservative indicators of concerns; e.g. "0" indicators *no* significant links with the CLASS. A less conservative approach would be to flag those standards that are only sometimes linked with the CLASS. ***Thus, in reading the table, the standards of greatest concern are those with greater numbers of 0s and xs.***

The large number of "0"s and "x"s in the table, especially for Learning and Development standards (LD), indicates that even though the overall/final QRIS ratings are modestly linked with CLASS scores, many of the standards themselves are either not linked with the CLASS, or are only inconsistently linked with the CLASS (e.g. for a specific CLASS domain, program type, or age group/CLASS tool). This is particularly the case for the Emotional and Organizational domains of the CLASS; more standards are linked with the Instructional domain in at least some instances. The large number of "x"s show that some standards (LD1, LD2, LD9, LD7, LD11, PQ1, PQ2) may be barriers that prevent higher quality programs from achieving a star-Level 3 or higher.

The largest number of concerns about standards (indicated by more "0" and "x" signifiers) were identified for Registered Family programs. This is likely a result of a) a smaller sample size of Registered Family programs; and b) less variability in QRIS ratings for Registered Family programs (most were levels 1 through 3). For example, as shown in Section 5 of this report, the size of the correlations between PQ1

and CLASS scores for Registered Family are as large as those for Centers, yet they do not reach statistical significance, likely due to sample size. This was also the case for LD8 and LD10 (see Appendix E). It is also possible that some QRIS standards and/or the CLASS do not measure quality as well for Registered Family programs as for other types of programs.

In addition to the overall pattern of findings, such as a fairly large number of "0"s and "x"s in the table overall, Table 12 also reveals patterns for specific standards. For example, while some standards, like LD5 and LD10 have few "0"s or "x"s, (indicating few concerns in their relationships with observed quality), others such as LD9 and LD11 (and also LD2, LD12, LD7, etc.) have many "0"s and "x"s. The standards in this second set are not well-linked with observed quality. For example, there are substantial concerns about LD9 (screening and assessment) across all three types of programs, and only limited evidence for significant links between LD9 ratings and CLASS scores for Registered Family providers (see Appendix E). Also, although LD11 (adult-child interactions) is conceptually very well-aligned with observed quality, only the 5-star indicator (observations) of LD11 ratings are significantly linked with CLASS scores. LD11 indicators at star-levels 3 and 4 were not linked with observed quality; these indicators involved written guidelines related to adult-child interactions. This lack of alignment of LD11 3-star and 4-star indicators with observed quality is particularly concerning given that many programs got stuck at the 2-star level on LD11, even many high quality programs. Thus, the written descriptions required for QRIS portfolio ratings in this domain appear to be largely unrelated to actual observed quality of interactions. The exception to this was among Registered Family programs, where there appears to be a link between the various LD11 ratings (3-star and higher) and CLASS scores, especially for Instructional Support. See Appendix E for more information about these, and other standards.

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

In Centers and Certified Family programs, in which the person writing guidelines is often someone other than the one(s) interacting with the children during the CLASS observation, there is no link between written guidelines and observed adult-child interactions.

The use of red coloring in Table 12 indicates that the QRIS Validation Study team suggests eliminating or substantially revising LD9, LD11, and LD12, due to the concerns described above. Orange coloring indicates additional standards (LD1, LD4, and LD6) that should be considered as candidates for elimination or revision. Blue is used to show an opportunity to strengthen and reduce redundancies of LD2 and LD7 by combining them. Based on the request of the mini review team, the validation study conducted supplemental analyses to explore the possibility of combining LD2 (curricula) and LD7 (planned curricular activities) into one new standard. Results suggest that such an approach would not only reduce the number of standards but would also strengthen the associations between these standards and observed quality (see Appendix E for more details).

Table 12. Areas of concern in how LD and PQ standards relate to CLASS scores.

	ALL			Centers			Certified Family			Registered Family*		
	ES	IS	OS	ES	IS	OS	ES	IS	OS	ES	IS	OS
LD 1 Philosophy	0x			0x			0	0	0	0x		
LD2 Curriculum Use	0		0	0		0	0		0	0x	0	0x
LD 3 Indoor Environment				0		0	0			0	0	
LD 4 Indoor Furnishings	0			0		0	0	0	0	0	0	0
LD 5 Outdoor Environment										0	0	0
LD 6 Materials	0		0	0		0		0		0x	0	0
LD 7 Planned Curricular Activities	0		0	0		0	0x	0		0x	0x	0x
LD 8 Routines						0		0		0	*0**	*0**
LD 9 Screening and Assessment	x	x	0x	0x	0x	0x	0x	0x	0x	x	x	x
LD 10 Group size/ratio/staffing						0				0	*0**	*0**
LD 11 Adult-Child Interactions	0x	x	x	0x		0x	0x	0x	0x	0x	x	0x
LD 12 Supports Social-Emot. Dev.	0	0		0	0	0	0	0	0	0		
PQ 1 Leader Qualifications							0	0		0x	0*x	0
PQ2 Teacher Qualifications	n/a	n/a	n/a	0x			n/a	n/a	n/a	n/a	n/a	n/a
PQ3 Assistant/Aide Qualifications	n/a	n/a	n/a					0		n/a	n/a	n/a
PQ4 Training							0	0			0	0

0 = no significant differences between star ratings for any CLASS tool (preK, toddler, combined, or total) and no significant correlations
x = more than 20% of programs with a rating of "2" scored high on this CLASS domain
* small sample size for Registered Family appears to limit significance of links with CLASS scores; sizes of the correlations are similar to other program types.
Colors are used to denote suggested revisions.

Red = substantial concern; suggest elimination or revision.
Orange = candidate for revision or elimination.
Blue = suggest combining LD2 and LD7 to strengthen and reduce redundancy.

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

Accurate and valid measures of the qualifications and training of program staff are critical in a QRIS. In Oregon, the existence of personnel measures beyond QRIS ratings make it possible to increase understanding of the validity of QRIS ratings within the Personnel Qualifications domain. This is particularly advantageous given the limitation, noted in Results Section 2, that the QRIS domains and standards are so highly correlated with one another that it is not possible to isolate the domains and standards most associated with higher levels of observed quality.

By accessing two additional sets of personnel measures 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. Each of the three sets of personnel measures (QRIS ratings and two additional sources) relied on the Oregon Registry Online (ORO) database, but each was created independently. ORO contained education and training data on persons employed in regulated child care facilities. Each person was linked to the facility in which she was employed. The three separate sets of personnel measures created from the ORO data were:

- QRIS ratings on personnel (PQ standards) for programs that earned a final QRIS rating of 2-5 based on steps on Oregon Registry as well as training hours. The range of correlations among PQ standards was $r = .60$ to $r = .71$ (correlation tables are available in Appendix E).
- Validation Study ORO (VS-ORO) measures for programs with final QRIS ratings of 2-5 based on steps on the Oregon Registry and training hours. The Validation Study team created multiple measures of personnel qualifications. For this analysis we used a single measure of each PQ construct so that correlations would be comparable with those from the PQ measures. The range of correlations among VS-ORO personnel measures ranged from $r = .02$ to $r = .46$ (correlation tables are available in Appendix E).
- Structural Indicators (SI) for programs with final QRIS ratings of 1-5 based on steps on the Oregon Registry, training hours, and education. These measures have been created annually since 2010. The 2014 SI measures were used for this analysis. Given the large number of SI variables, we reported the correlations by program type. For this analysis we used a single measure of each PQ construct so that correlations would be comparable with those from the PQ measures. The range of correlations among SI personnel measures ranged from $r = .03$ to $r = .49$ for centers, $r = .00$ to $r = .46$ for CF programs, and $r = .06$ to $r = .44$ for RF providers (correlation tables are available in Appendix E).

Examining each set of personnel measures separately, we found that the QRIS PQ measures more highly correlated with one another than were either the VS-ORO or SI personnel measures, which re-affirmed that the QRIS rating process led to artificially high correlations among standards with the QRIS than would otherwise occur (see Results Section 2 for further explanation). A fuller description of each measure is found in Appendix E, Section 5.

Our research questions for this analysis focused on the extent to which these different measures of personnel qualifications and training were correlated with observed quality (CLASS scores) and final QRIS star ratings. Before addressing those questions, we first examined how correlated the three sets of personnel measures were with each other.

It is important to note that the SI dataset included Level 1 programs whereas the QRIS and VS-ORO data included only programs with a final star rating of 2 to 5. Thus, there may have been more measurable variability in the analyses conducted with the SI dataset. Further, identification of Level 1 programs was based on SI data. These two factors increased the likelihood of finding significant correlations between SI measures and CLASS scores.

Correlations among the Three Sets of Personnel Measures

We examined correlations among different measures of personnel qualifications by four key QRIS PQ constructs:

- PQ1: Director/provider qualifications (Registry step and/or education level),
- PQ2: Center teacher qualifications (Registry step and/or education level),
- PQ3: Center aide/Certified Family assistant qualifications (Registry step and/or education level), and
- PQ4: Staff training (number of hours per year).

We found that, except for training, the three sets of personnel measures were moderately to highly correlated with each other (r = .26 to r = .73). The education measures existed only in the SI dataset and were only sometimes correlated with other measures (r = .04 to r = .64). Correlations among the three sets of personnel measures other than education were highest for director/provider (r = .30 to r = .73) and teacher (r = .46 to r = .72) measures. For director/provider they were higher for Registered and Certified Family providers than for Centers.

> The three sets of personnel measures were moderately to highly correlated with each other.

The training measures were the least correlated (r = .04 to r =. 58). Given that all three sets of personnel measures were created from the same raw data (ORO), this finding demonstrated that how a measure was created or operationalized mattered and pointed to the need to carefully craft and define measures to most accurately capture personnel qualifications and training, especially training. See Appendix E, Section 5, for detailed description of the three sets of personnel measures, more detail on correlations, and correlation tables.

Correlations among Personnel Measures and Observed Quality (CLASS Scores)

Having found that personnel measures were moderately to highly correlated, we then addressed the question, "How correlated are the three sets of measures with observed quality as measured by the CLASS?" In this analysis we brought together key measures from each of the three sets of personnel measures and examined their correlations with CLASS scores. We selected personnel measures that prior analyses indicated were likely to be correlated with observed quality. Given that we found substantial differences in personnel measures by type of care, we examined the correlations for each care type separately.

Centers

Table 13 shows the correlations between Center personnel measures and observed quality. **For Centers we find modest correlations between a few personnel measures and CLASS scores.** For director qualifications, both the QRIS (PQ1) and VS-ORO measures are modestly correlated with the Organized Classrooms domain. In contrast, the SI measure of director qualifications is linked with the Instructional Support domain. This pattern of findings across the three sources increases confidence that there is a

link between the qualifications of directors and the quality of the adult-child interactions that take place within Centers, although it is not completely clear whether this is most apparent for the Organized Classrooms or the Instructional Support domains. The only source of teacher qualifications measures that is significantly linked with any of the CLASS domains in Centers is the SI measure, which is linked with Instructional Support. This may be due to the inclusion of the Level 1 programs in the SI data but we cannot be certain. No links are found between aide qualifications and CLASS scores. The only training hours measure linked with CLASS scores is the QRIS (PQ4) rating. It is possible that this association is due to the high correlations among the standards/domains within the QRIS; it may not reflect a real association between training hours and CLASS scores in Centers.

Table 13. Correlations among Personnel Measures and Observed Quality in Centers.

		CLASS Domains		
Construct	Measure	Emotional Support	Instructional Support	Organized Classrooms
Director	PQ1: Dir/prov qualifications	.15⁺	.13	**.23***
	VS-ORO director Registry step	-.04	.13	**.23***
	SI directors Registry step 9 or higher	.14⁺	**.20***	.03
	SI director has a degree	-.03	.06	-.03
Teacher	PQ2: teacher qualifications	.03	.13	.13
	VS-ORO teacher median step	-.05	.10	.10
	SI % teachers Registry step 9 or higher	.01	**.19***	-.04
	SI % teachers have a degree	.01	**.20***	.16
Aide/Assistant	PQ3: Aide/asst qualifications	.11	.11	.10
	VS-ORO aide median step	**.29***	-.22	-.01
	SI % aides Registry step 5 or higher	.21⁺	-.06	.02
	SI % aides have a degree	-.14	-.10	.12
Training	PQ4: Training	.06	.05	**.20***
	VS-ORO % staff 24 hours or more of training	-.13	-.07	.03
	SI % staff 20 hour or more of training	.15⁺	.11	.02

Notes: SI dataset includes programs level 1-5 whereas QRIS PQ and VS-ORO include only programs levels 2-5. Although both VS-ORO and SI contain multiple measures of each construct, we are presenting only one for simplicity/readability. We selected one that prior analysis indicated would be associated with CLASS scores. Degree includes an Associate as well as Bachelors or higher.
N-value varies for each correlation based on how many were observed. For QRIS and VS-ORO Minimum N=81 and Maximum N=120. For SI Minimum M=75 and Maximum N=140.
⁺ Correlation is nearing significance at the $p < .10$ level (2-tailed)
* Correlation is significant at the $p < .05$ level (2-tailed)

Certified Family Programs

Table 14 shows the correlations between Certified Family (CF) personnel measures and observed quality. **In the case of CF programs, personnel measures are more consistently correlated with observed quality than is the case with Centers; this is particularly the case for the Organized Classrooms domain but is also notable for the Emotional and Instructional domains.** These findings

provide confidence that provider qualifications are associated with observed quality for Certified Family programs. Assistant qualifications are only slightly less consistently associated, as a number of measures are associated with Organized Classroom scores and one SI measure is moderately associated with Instructional Support scores. Similarly, findings provide confidence that training of Certified Family staff is associated with observed quality; training measures from the three different datasets are associated with at least one CLASS domain.

Table 14. Correlations among Personnel Measures and Observed Quality in Certified Family.

		CLASS Domains		
Source	Measure	Emotional Support	Instructional Support	Organized Classrooms
Director	PQ1	.16	.21^{+}	**.35****
	VS-ORO Registry Step 9 or higher	.16	**.27***	**.28***
	SI Registry Step 9 or higher	**.23***	**.27***	**.41****
	SI Provider has a Degree	**.23***	.05	**.30***
Assistant	PQ3	.16	.16	**.39****
	VS-ORO assistant median Registry step	.17	.19	**.37***
	SI % assistants Registry step 5 or higher	.17	**.44*****	**.45****
	SI % assistants have a degree	.13	.22	.14
Training	PQ4	.11	.19^{+}	**.32***
	VS-ORO % staff 24 hours or more of training	.04	**.26***	.24^{+}
	SI % staff 20 hours or more of training	**.30****	.19	.16

Notes: SI dataset includes programs level 1-5 whereas QRIS PQ and VS-ORO include only programs levels 2-5. Although both VS-ORO and SI contain multiple measures of each construct, we are presenting only one for simplicity/readability. We selected one that prior analysis indicated would be associated with CLASS scores. Degree includes an Associate as well as Bachelors or higher.
N-value varies for each correlation based on how many were observed. Minimum N=44 and Maximum N=81
$^{+}$ Correlation is nearing significance at the $p < .10$ level (2-tailed)
* Correlation is significant at the $p < .05$ level (2-tailed)
** Correlation is significant at the $p < .01$ level (2-tailed)
*** Correlation is significant at the $p < .001$ level (2-tailed)

Registered Family Programs

Table 15 shows the correlations between Registered Family personnel measures and observed quality. **For Registered Family programs, fewer personnel measures are correlated with observed quality than is the case for the other two types of care. This may be due, in part, to limitations such as a smaller sample size, less variation in QRIS ratings (most are levels 1-3), and fewer potential personnel measures to test in association with observed quality for Registered Family than for the other two types of care. Each of these differences reduces the likelihood of finding significant correlations.**

The pattern of findings for the qualifications of Registered Family programs (Table 15) is suggestive of a possible association with observed quality. The size of the correlations ($r = .22$ to $r = .25$) between the

PQ and VS-ORO Registered Family provider qualifications and Instructional Support are similar to those that are statistically significant for Centers (see Table 13). The lack of significance for Registered Family is likely due to limited power related to the small sample size. This is supported by the finding that the SI provider qualification measure is trending toward significance in its association with instructional Support scores.

For Registered Family provider training, the VS-ORO measure of training is associated significantly with both Emotional Support and Organized Classroom scores, and is trending toward significance in its association with Instructional Support. The SI training measure is moderately associated with Organized Classroom scores. These findings indicate that training makes a difference for these providers of small home-based care.

Table 15. Correlations among Personnel Measures and Observed Quality in Registered Family.

		CLASS Domains		
Construct	**Measure**	**Emotional Support**	**Instructional Support**	**Organized Classrooms**
Provider Qual	PQ1	-.07	.25	.19
	VS-ORO Provider Registry step	-.07	.22	.12
	SI Prov Registry step 8 or higher	-.05	.24^{+}	.00
	SI Provider has a degree	-.07	.08	-.09
Training	PQ4	.08	.19	.18
	VS-ORO staff has 18 hours or more of training	**.33***	.26^{+}	**.34***
	SI provider has 20 hours or more of training	.01	.18	**.35***

Notes: SI dataset includes programs level 1-5 whereas QRIS PQ and VS-ORO include only programs levels 2-5. Although both VS-ORO and SI contain multiple measures of each construct, we are presenting only one for simplicity/readability. We selected one that prior analysis indicated would be associated with CLASS scores. Degree includes an Associate as well as Bachelors or higher.
N-value varies for each correlation based on how many were observed. Minimum N=33 and Maximum N=59
$^{+}$ Correlation is nearing significance at the $p < .10$ level (2-tailed)
* Correlation is significant at the $p < .05$ level (2-tailed)

A Summary of Correlations among Personnel Measures and Observed Quality

As noted earlier in this report, the high levels of correlations between and within QRIS domains limit the Validation Study team's ability to examine which domains or standards are associated with observed quality. Having three distinct sets of measures provides the opportunity to broaden understanding of how personnel measures are correlated with observed quality. Finding that a sizeable number of personnel measures are correlated with one or more domains of the CLASS, we entered these findings into a single table to more effectively display what we have learned (Table 16). A number of insights into personnel measures emerge:

- The three sets of personnel measures are associated with at least some of the domains of observed quality—suggesting that personnel qualifications and training are associated with observed quality.

- Although we find associations between personnel measures and CLASS scores, findings are not consistent. We find differences across type of care and across different CLASS domains. This means that slight differences in the way that personnel measures are structured can change the way the measures relate to observed quality. We can be most confident in associations when they are consistent, such as is the case for Certified Family programs.
- The qualifications of the program leader appear to be linked with the quality of adult-child interactions. All three distinct measures of **Center director and Certified Family provider qualifications** are correlated with one or CLASS domains. **Registered Family provider** qualifications appear likely to be associated with Instructional Support scores if the sample had been larger.
- For **Center teacher qualifications and aide qualifications**, some of the VS-ORO and SI measures are correlated with observed quality whereas among QRIS PQ measures only PQ3 for Certified Family assistants is correlated with observed quality.
- All three distinct measures of **staff training** are correlated with observed quality for at least one type of care; PQ4 for Centers and Certified Family, VS-ORO for both Certified Family and Registered Family, and SI for Certified Family and Registered Family and trending toward significance for Centers.
- The evidence of the association between personnel measures and observed quality is the strongest for Certified Family programs. The only personnel measures not correlated with one or more observed quality scores for Certified Family programs is whether or not the assistants have a degree.

Table 16. Summary of Correlations among Personnel Measures and Observed Quality

Construct	Personnel Measure	Type of Care		
		Centers	CF Programs	RF Programs
Director/ Provider Qualifications	PQ1	ES+, OC	IS+, OC	--
	VS-ORO director Registry step	OC	IS, OC	--
	SI director Registry Step 9 or higher	IS	ES, IS, OC	IS+
	SI director some college or degree	--	ES, IS, OC	--
Center Teacher Qualifications	PQ2	--	NA	NA
	VS-ORO median Registry step	--	NA	NA
	SI-% teachers Registry step 9 or higher	IS	NA	NA
	SI % teachers some college or degree	IS	NA	NA
Center Aide CF Assist Qualifications	PQ3	--	OC	NA
	VS-ORO aide/assistant median Registry step	ES	OC	NA
	SI % aides/assistants Registry step 5 or higher	ES+	IS, OC	NA
	SI % aides/assistants ECE degree	-OC+	--	NA
Staff Training	PQ4	OC	IS+, OC	--
	VS-ORO % staff training 24 hours or more (18 for RF)	--	IS, OC+	ES, IS+, OC
	SI % staff training 20 hours or more	ES+	ES	OC

Notes: Entries in this table represent domains of the CLASS for which a statistically significant ($p < .05$) correlation was detected. ES = Emotional Support. IS = instructional Support; OC = Organized Classrooms. + Correlation is nearing significance at the $p < .10$ level (2-tailed).
SI dataset includes programs level 1-5 whereas QRIS PQ and VS-ORO include only programs levels 2-5. Although both VS-ORO and SI contain multiple measures of each construct, we are presenting only one for simplicity/readability. We selected one that prior analysis indicated would be associated with CLASS scores. Degree includes an Associate as well as Bachelors or higher.
N-value varies for each correlation based on how many were observed. For Centers: Minimum N=57 and Maximum N=140. For CF: Minimum N=45 and Maximum N=81. For RF: Minimum N=35 and Maximum N=58.

Correlations of Personnel Measures with Final Star Rating

Next we examined the association between personnel measures and final star rating to examine whether personnel measures affected a program's final rating. Also, given availability of SI measures for all regulated programs in Oregon, finding that the SI measures were associated with final star rating would increase confidence that Oregon has information on the quality of all regulated programs.

Table 17 shows that significant correlations are modest to high across all types of care and for many of the measures ($r = .19$ to $r = .76$) but that Certified Family and VS-ORO measures for Center aides as well as the SI education and VS-ORO training for Certified Family assistants are not. As expected,

correlations for most QRIS PQ measures are higher than are those for VS-ORO and SI measures, likely due to the high correlations among QRIS standards and domains (see Results Section 2).

Table 17. Summary of Correlations among Personnel Measures and Final QRIS Star Rating

Construct	Personnel Measure	Type of Care		
		Centers	CF Programs	RF Programs
Director/ Provider Qualifications	PQ1	**.70*** **	**.76*** **	**.72*** **
	VS-ORO director Step	**.37*** **	**.70*** **	**.53*** **
	SI director Registry Step 9 or higher (8 for RF)	**.38*** **	**.67*** **	**.62*** **
	SI director degree	**.35*** **	**.38*** **	**.23**+**
Center Teacher Qualifications	PQ2	**.65*** **	NA	NA
	VS-ORO teacher median Registry step	**.54*** **	NA	NA
	SI-Registry step 9 or higher	**.56*** **	NA	NA
	SI teacher degree	**.38*** **	NA	NA
Center Aide CF Assist Qualifications	PQ3	**.62*** **	**.59*** **	NA
	VS-ORO median Registry step	.17	**.43** **	NA
	SI aide/asst Registry step 5 or higher	.18	**.46*** **	NA
	SI aide/asst some college or degree	-.04	.22	NA
Staff Training	PQ4	**.63*** **	**.70*** **	**.53*** **
	VS-ORO staff training 24 plus hours (18 for RF)	**.19***	.16	**.56*** **
	SI staff training 20 plus hours	**.61*** **	**.31** **	**.68*** **

Notes: SI dataset includes programs level 1-5 whereas QRIS PQ and VS-ORO include only programs levels 2-5. Although both VS-ORO and SI contain multiple measures of each construct, we are presenting only one for simplicity/readability. We selected one that prior analysis indicated would be associated with CLASS scores. Degree includes an Associate as well as Bachelors or higher.
N-value varies for each correlation based on how many were observed. For Centers: Minimum N=57 and Maximum N=140. For CF: Minimum N=45 and Maximum N=81. For RF: Minimum N=35 and Maximum N=58.
+ Correlation is nearing significance at the $p < .10$ level (2-tailed)
* Correlation is significant at the $p < .05$ level (2-tailed)
** Correlation is significant at the $p < .01$ level (2-tailed)
*** Correlation is significant at the $p < .001$ level (2-tailed)

A Summary of Correlations among Personnel Measures and QRIS Final Star Rating

Insights from the exploration of correlations among personnel measures, CLASS scores, and final star ratings include:

- QRIS PQ ratings are more highly correlated with final ratings than are VS-ORO and SI measures, especially for Centers. This is likely the result of the QRIS design that results in individual domain and standard ratings moving together rather than independently.

- The majority of personnel measures are moderately to highly correlated with the programs' QRIS final star rating (r =. 31 to r = .70), with the exception of the VS-ORO and SI measures for aide/assistant which are not significant and small (r =. 04 to r =. 18).
- Since VS-ORO and SI personnel measures are created outside the QRIS rating process we have confidence that the associations are not affected by QRIS ratings from other domains. Finding associations between these measures and final star rating increases confidence that personnel qualifications and training are associated with final star ratings.
- Most measures for Certified Family programs are more highly correlated with final star ratings than are measures for Centers and Registered Family programs. Capturing personnel measures in Centers appears more complicated than for Family programs, possibly due to the larger numbers of personnel in Centers.

Summary of Findings on Additional Structural Indicator Measures

The Structural Indicator dataset contained an additional four measures that research suggested might be related to quality: teacher wages, teacher benefits, retention, and accreditation. Wages and benefits were only applicable for Centers but retention and accreditation were meaningful for all types of care. We examined the associations between these measures and both observed quality (see Table 18) and final star rating (see Table 19). The lowest wage a Center paid teachers was associated with Instructional Support scores in Centers and was trending toward significance with Organized Classroom scores. Benefits were also included in QRIS but the QRIS and SI benefits measures were not correlated with one another, providing evidence of the challenge of accurately capturing benefit practices. Provider retention was associated with Organized Classroom scores for Registered Family providers. Center teacher wages were also associated with final star rating and accreditation was associated with final star ratings of Centers and Certified Family programs; this may not have been correlated for Registered Family programs because of the small number of them that are accredited.

The association of Center teacher wages with both observed quality and final star rating seems to indicate that a wage measure is worth further consideration. Finding that retention of Registered Family providers is associated with observed quality seems to indicate that a retention measure for Family providers is worth further consideration. Neither benefits nor accreditation are associated with observed quality, and accreditation's association with final star rating could be due to QRIS fast tracking of accredited programs.

Table 18. Summary of Correlations among Personnel Measures and Observed Quality

Construct	Personnel Measure	Type of Care		
		Centers	CF Programs	RF Programs
Teacher Wages	Center teacher lowest wage	IS, OC^{+}	NA	NA
Benefits	Benefits	--	NA	NA
Retention	Teacher/ provider retention	--	--	OC
Accreditation	Accreditation	--	--	--

Notes: Entries in this table represent domains of the CLASS for which a statistically significant ($p < .05$) correlation was detected. ES = Emotional Support. IS = instructional Support; OC = Organized Classrooms. $^{+}$ Correlation is nearing significance at the $p < .10$ level (2-tailed).
SI dataset includes programs level 1-5.
N-value varies for each correlation based on how many were observed. For Centers: Minimum N=75 and Maximum N=141. For CF: Minimum N=62 and Maximum N=77. For RF: Minimum N=45 and Maximum N=59.

Table 19. Summary of Correlations among Personnel Measures and Final QRIS Star Rating

Construct	Personnel Measure	Type of Care		
		Centers	CF Programs	RF Programs
Teacher Wages	Center teacher lowest wage	**.38*** **	NA	NA
Benefits	Benefits	.05	NA	NA
Retention	Teacher/ provider retention	.16^{+}	.09	-.03
Accreditation	Accreditation	**.41*** **	**.34** **	.08

Notes: SI dataset includes programs level 1-5.
N-value varies for each correlation based on how many were observed. For Centers: Minimum N=82 and Maximum N=141. For CF: N=77. For RF: N=59.
$^{+}$ Correlation is nearing significance at the $p < .10$ level (2-tailed)
* Correlation is significant at the $p < .05$ level (2-tailed)
** Correlation is significant at the $p < .01$ level (2-tailed)
*** Correlation is significant at the $p < .001$ level (2-tailed)

Concluding Thoughts

VS-ORO and SI personnel measures provide the opportunity to explore the association between personnel qualifications/training and both observed quality and final star rating without concern that the ratings are being affected by ratings from other domains and standards.

For Centers, the personnel measures most closely linked with observed quality are: director registry step, teachers having either step 9 or higher, or a degree, and the median step for assistants. Training is not linked with observed quality in Centers. For Certified Family programs, the personnel measures most linked with observed quality are 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 is staff training. The associations between the providers' registry step and the CLASS are suggestive of a possible relationship. They are similar in size to those for Centers but are not statistically significant, likely due to limited statistical power from a small sample size.

Findings indicate that measuring training is challenging, and that capturing personnel qualifications and training in Centers is particularly difficult, possibly due to the larger numbers of personnel in Centers.

The association between personnel measures created independently from the QRIS PQ ratings and QRIS final star rating indicates a fairly strong link between the qualifications and training of personnel in a program and the final star rating a program achieves. These additional personnel measures are at least as consistently linked with CLASS scores as the PQ ratings, and often more so. This increases confidence that personnel qualifications (for Centers and Certified Family; possibly for Registered Family) and training (for Certified and Registered Family) are linked with observed quality. It also points to the need to strengthen the personnel qualification and training measures used in QRIS.

Additionally, the Structural Indicators provide data on all registered programs statewide. Findings from this study indicate that these data relate to the quality of adult-child interactions in a meaningful way and thus provide some level of information on the quality of all regulated programs.

Summary and Conclusions

This final section of the report summarizes key findings from each of the five research questions, and highlights considerations and implications for the future of Oregon's QRIS.

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

QRIS Ratings. Of the 304 programs participating in the QRIS Validation Study, 19% were Level 1 programs and 81% were rated by the QRIS at levels 2 through 5. The Level 1 programs were recruited as a "low quality" comparison group of programs not participating in the QRIS. 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%).

Since programs must pass all 5 domains (Learning and Development (LD), Personnel Qualifications (PQ), Health and Safety (HS), Family Partnerships (FP), and Administration and Business Practices(AB)) to achieve a star-level, programs' QRIS ratings are based on the domain for which they rated the *lowest*. Further, programs must pass most or all of the standards within a given domain in order to achieve the targeted star-level. Certain standards were much harder for programs than others. Standards with more than one-quarter of programs scoring a 2 were LD9, LD11, and HS6, although several other standards also have relatively high percentages of programs scoring a 2. For Registered Family providers, there were additional standards for which more than one in four of them scored a 2: LD1, LD7, HS1, HS3, HS6, PQ1, and AB5.

CLASS scores. Observations of adult-child interactions using the Class were scored on a 7-point scale from 1 (very low) to 7 (very high). 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.

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. This may be because many of the QRIS standards require formal policies, written procedures, specific types of furnishings and materials, etc., whereas the observed quality measure (CLASS) deals specifically with how adults interact with young children.

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

This study found that the five domains of Oregon's QRIS are highly correlated, as are the standards within each domain. These correlations are much larger than those from prior studies of similar constructs in the field (e.g. Burchinal et al., 2016). They appear artificially high; likely a result of the

portfolio/block structure of Oregon's QRIS. These high correlations among the various parts of the QRIS 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 the 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 to consider 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, such as those created using Oregon Registry Online data (VS-ORO) or Structural Indicators (SI; see Section 5 for more about ORO and SI).

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 rated 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 partially related to lower observed quality in Level 1 programs; differences were smaller and less consistent when only comparing programs rated 2 vs 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.

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.

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.

There are several possible reasons that the associations 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 limits the strength of associations between programs' QRIS ratings and observed quality.

- The differences between higher- and lower-quality programs were small. Few programs provided high quality care, as measured by the CLASS. 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 are not large enough to translate into large associations between QRIS ratings and observed quality.

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

Findings from this 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.

Even though the overall QRIS ratings are linked somewhat with CLASS scores, many of the standards themselves are either not linked with the CLASS, or are only inconsistently linked with the CLASS (e.g. for a specific CLASS domain, program type, or age group/CLASS tool). This is particularly the case for the Emotional and Organizational domains of the CLASS; more standards are linked with the Instructional domain in at least some instances. Fewer standards were associated with CLASS scores for Registered Family programs.

An example of a standard that was identified as concerning was LD9 standard (screening and assessment). LD9 was not only very difficult for providers but was also not linked with observed quality. In another example, LD11 (adult-child interactions) is conceptually very well-aligned with observed quality, but only the 5-star indicator (observations) of LD11 ratings are significantly linked with CLASS scores. LD11 indicators at star-levels 3 and 4 were not linked with observed quality; these indicators involved written guidelines related to adult-child interactions. Relying on written guidelines may not be an appropriate or valid indicator of the quality of adult-child interactions.

These types of concerns may be important to consider, alongside other sources of information, in efforts to strengthen Oregon's QRIS. Findings revealed substantial concerns regarding LD9, 11, and 12; 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 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. Each of the three sets of personnel measures (QRIS ratings and two additional sources) relied on the Oregon Registry Online (ORO) database, but each was created independently.

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. Training was not linked with observed quality in Centers. 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' registry step and the CLASS were suggestive of a possible relationship. There were similar in size to those for Centers but were not statistically significant, likely due to limited statistical power from the small sample size.

Slight variations in how variables were constructed often led to differences in their associations with observed quality.

Careful attention must be paid to how to utilize the ORO data as indicators of quality.

Slight variations in how variables were constructed from the ORO database often led to differences in their associations with observed quality (see Section 5 and Appendix E for more information). Careful attention must be paid to how to utilize the ORO data as indicators of quality.

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

Additionally, the majority of personnel measures were moderately to highly correlated with the programs' QRIS final star rating. This indicates a fairly strong link between the qualifications and training of the personnel in a program and the final star rating that program achieves.

These additional personnel measures were at least as consistently linked with CLASS scores as the PQ ratings, and often more so. This increases confidence that personnel qualifications (for Centers and Certified Family; possibly for Registered Family) and training (for Certified and Registered Family) are linked with observed quality. These measures should be considered as possible replacements for the current PQ standards. It will be critical, however, that personnel measures remain intuitive and understandable to 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.

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

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. These Structural Indicators could be more directly built into Oregon's QRIS, which is intended to apply to all regulated programs in the State.

Strengths and Limitations of the Study Design

This study had several methodological strengths that contribute to confidence in findings and to utility of the results. Programs from all three types of regulated care (Registered Family, Certified Family, and Certified Centers), and from across the State of Oregon participated. This means that the results are representative of the breadth of programs in Oregon. The inclusion of Level 1 programs that were not participating in the QRIS and had low levels of personnel qualifications/training provided a "low quality" comparison group, and increased the variability in quality of programs in the study. This increased our ability to detect differences in observed quality between programs rated 3-star or higher and those who did not. Additionally, the use of multiple measures of personnel led to increased confidence that personnel qualifications/training are linked with observed quality of adult-child interactions. Finally, the analyses involved multiple approaches to triangulate evidence, increasing confidence in the findings. The deep dive into exploration of specific standards and their associations with observed quality provides insight into concrete ways to strengthen the QRIS.

As with any study, limitations also hindered our ability to draw conclusions from the data. In particular, the relatively small sample of Registered Family providers led to limited variability and less statistical power than for the other two program types. Thus, some of the non-significance in associations with observed quality are likely due to small sample size but it is impossible to know for sure that this is the case.

Additionally, the limited research literature on the Toddler CLASS makes it difficult to ascertain the reasons for the lack of associations between programs' QRIS ratings and their scores on the Toddler CLASS. We cannot know whether the Toddler CLASS instrument was not as valid of an instrument, and therefore did not "work well", or whether Oregon's QRIS standards were not as applicable to quality of toddler-aged classrooms and therefore did not differentiate quality on the Toddler CLASS. It is possible that programs met standards that are based on one classroom (as with LD10 regarding group size/ratio), or a percentage of staff having certain qualifications (as with PQ2 and 3) with their preschool-aged classrooms more than with their toddler-aged classrooms. The study design does not allow us to determine this.

Next Steps for the Validation Study

Oregon's QRIS Validation Study is currently conducting observations of child and family engagement in programs across the State. A report on associations between the QRIS and these additional outcomes is forthcoming. Such findings promise to providing additional information relevant to strengthening Oregon's QRIS and ultimately to improving outcomes for children and families.

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 rating system somewhat differentiates the quality of the interactions that young children have with the adults that care for them in regulated programs across the state. Yet, differences tended to be small in size, and only apparent when contrasting programs rated 3-star or higher to those at level 1 or 2. We did not find consistent evidence that programs rated 4- or 5-star provide 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. Instructional Support involves rich conversations and back and forth exchanges that encourage children to think deeply and strengthen language/literacy skills. Programs provided Emotional Support (emotional climate, sensitivity, regard for student perspectives) that was consistently in the upper end of the "mid" range. Programs rated higher on the QRIS did not provide higher Emotional Support. Scores for Organized Classrooms tended to be in the mid-range, although Certified, and sometimes Registered Family programs rated 3-star or higher were sometimes higher than Organizational scores for programs rated 1 and 2. The Organized Classrooms domain focuses on behavior management, productivity, and learning formats. Children who receive higher quality care in these three domains, especially in Instructional Support, during the preschool years show stronger school readiness skills upon entry to elementary school (e.g. Hamre, Hatfield, Pianta, & Jamil, 2014). How much of a difference in quality is enough to improve outcomes, however, remains unclear (e.g. Burchinal et al., 2010; Hatfield et al., 2016).

Children attending programs rated 3-star or higher appear to experience somewhat higher quality interactions with their caregivers than those attending level 1 or 2 programs.

Whether this represents a large enough difference to translate into better outcomes for children remains unknown.

In other words, children attending programs rated 3-star or higher appear to experience somewhat higher quality interactions with their caregivers than those attending level 1 or 2 programs, but whether this represents a large enough difference to translate into better outcomes for children remains unknown. Findings from studies of other QRISs across the country are mixed (e.g. Karoly, 2014). For example, a recent validation study of Minnesota's QRIS found ratings linked with only 2 out of 8 measures of children's development (Tout et al., 2016). Other studies have found no associations between QRIS ratings and children's outcomes (e.g. Magnusson & Lin, 2016; Soliday Hong, Howes, Marcella, Zucker, & Huang, 2015).

Study Two of Oregon's QRIS Validation Study is currently examining links between QRIS ratings and an observational measure of child engagement, as well as a parent-report measure of family engagement.

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

Findings point to a mix of similarities and differences in how Oregon's QRIS relates to observed quality by program type. The overall conclusion, that programs rated 3-star or higher provide somewhat higher quality care to young children 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 tend to be lower. Few achieved 4- or 5-star ratings. This discrepancy calls for revisions to Oregon's QRIS. One potential solution is to revise and/or eliminate standards that serve as barriers to higher ratings if they do not clearly represent differences in quality (see below, and Results Section 4). Additional solutions, such as targeted technical assistance and improved educational pathways, may also be possible but are beyond the scope of this study.

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. Furthermore, in-depth analysis of various personnel measures suggested that higher observed quality may only be linked with a high level of qualifications (step 9 on the Oregon Registry) for teachers in Centers. To strengthen the link between QRIS ratings and children's experiences in their actual classrooms the QRIS should consider increasing consistency in what is required across classrooms/groups/teachers. This increased rigor could be balanced by eliminating standards that are currently creating barriers to achieving ratings without relating to observed quality (see Results Section 4 and below).

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 did provide 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, 11, and 12; we suggest either eliminating or substantially revising these standards. Additional standards that should be considered as candidates for eliminating or revising include LD1, 4, and 6. Additionally, the Validation Study team found that the LD domain could be strengthened by combining LD2 and LD7 into one new standard. Collectively these changes have the potential to reduce the number of programs that provide higher quality care to children who fail to achieve a star-Level 3 or higher. This may be particularly important for Registered Family programs, few of which were able to achieve 4- and 5-star ratings.

We also have confidence that personnel qualifications and/or training are linked with observed quality, due to triangulating evidence across multiple sources of data apart from the QRIS ratings.

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.
- Consider streamlining other standards and domains in addition to LD and PQ 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/personnel in programs with multiple classrooms/groups. This increased rigor could be offset by eliminating standards that are currently creating barriers to achieving ratings without relating to observed quality (listed above).
- 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.

References

Build Initiative. (2015). *Current status of QRIS in States.* Retrieved on June 4, 2016 from http://qrisnetwork.org/sites/all/files/maps/QRISMap.pdf

Burchinal, M., Soliday Hong, S., Sabol, T., Forestieri, N., Peisner-Feinberg, E., Tarullo, L. & Zaslow, M. (2016). Quality Rating and Improvement Systems: Secondary data analyses of psychometric properties of scale development. OPRE Report #2016-26. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Child Trends and Build Initiative. (2014). *QRIS Online Compendium*. Retrieved on June 5, 2016 from http://qriscompendium.org/

Early, D., Barbarin, O., Bryant, D., Burchinal, M., Chang, F., Clifford, R., . . . Barnett, W.S. (2007). *Prekindergarten in eleven states: NCEDL's multi-state study of prek-kindergarten and study of state-wide early education programs.* Frank Porter Graham Child Development Institute, University of North Carolina at Chapel Hill.

Guy, S., & Aldrich, P. (April, 2015). Oregon Quality Rating and Improvement System Process Evaluation. FY2015. Retrieved from: http://triwou.org/projects/qris/resources

Hamre, B.K., & Pianta, R.C. (2007). Learning opportunities in preschool and early elementary classrooms. In R.C. Pianta, M.J. Cox, & K.L. Snow (Eds.), *School readiness and the transition to kindergarten in the era of accountability,* (pp.49-83). Baltimore: Paul H. Brookes Publishing.

Helburn, S.W. (Ed.) (1995). Cost, quality, and child outcomes. Retrieved from http://files.eric.ed.gov/fulltext/ED386297.pdf

Joseph, G.E., Feldman, E.N., Brennan, C., Naslund, R., Phillips, J., & Petras, A. (2011). *Seeds to Success field test: Year two – final technical report.* Childcare Quality & Early Learning, Center for Research and Training, University of

Washington. Available at: http://qrisnetwork.org/sites/all/files/resources/gscobb/2011-09-14%2006%3A11/WashingtonStateQRISPilotReport.pdf

Karoly, L. A. (2014). *Validation studies for early learning and care quality rating and improvement Systems: A review of the literature.* Rand Corporation. Retrieved on June 4, 2016 from http://www.rand.org/content/dam/rand/pubs/working_papers/WR1000/WR1051/RAND_WR1051.pdf

La Paro, K.M., Hamre, B., & Pianta, R.C. (2011). *Classroom Assessment Scoring System: Toddler manual.* Baltimore: Paul H. Brookes Publishing.

La Paro, K.M., Pianta, R.C., & Stuhlman, M. (2004). The Classroom Assessment Scoring System: Findings from the prekindergarten year. *The Elementary School Journal*, *104*(5), 409-426.

Laughlin, L. (2013, April) *Who's minding the kids: Child care arrangements Spring 2011.* Retrieved from https://www.census.gov/prod/2013pubs/p70-135.pdf

Morrison, F.J., & Connor, C.M. (2002). Understanding schooling effects on early literacy: A working research strategy. *Journal of School Psychology*, *40*(6), 493-500.

NICHD Early Child Care Research Network. (2003). The NICHD study of early child care: Contexts of development and developmental outcomes over the first seven years of life. In J. Brooks-Gunn, A. S. Fuligni, & L. J. Berlin (Eds.) *Early child development in the 21st century*. (pp. 181-201). New York: Teachers College Press.

Pianta, R.C., La Paro, K.M., Payne, C., Cox, M.J., & Bradley, R. (2002). The relation of kindergarten classroom environment to teacher, family, and school characteristics and child outcomes. *The Elementary School Journal, 102*(3), 225-238.

Pianta, R.C., La Paro, K.M., & Hamre, B.K. (2008). *Classroom Assessment Scoring System: PreK manual.* Baltimore: Paul H. Brookes Publishing.

Rieber, R.W. (1998). *The collected works of L.S. Vygotsky: Vol. 5 Child Psychology.* New York: Plenum.

Rutter, M., & Maughan, B. (2002). School effectiveness findings, 1979-2002. *Journal of School Psychology,* 40(6), 451-475.

Vandell, D. L., & Wolfe, B. (2000). Child care quality: Does it matter and does it need to be improved? Retrieved from http://www.irp.wisc.edu/publications/sr/pdfs/sr78.pdf

Weber, R. B., & Wolfe, J. (2003). Improving child care: Providing comparative information on child care facilities to parents and the community. (Child Care Policy Research Issue Brief). Retrieved from http://health.oregonstate.edu/sites/default/files/sbhs/pdf/2003-Improving-ChildCare.pdf

Hamre, B. K., Hatfield, B. E. , Pianta, R. C., & Jamil, F. (2014). Evidence for General and Domain-Specific Elements of Teacher–Child Interactions: Associations With Preschool Children's Development. *Child Development*, 85(3), 1257-1274. doi:10.1111/cdev.12184

Hatfield, B. E., Burchinal, M. R., Pianta, R. C., & Sideris, J. (2016). Thresholds in the association between quality of teacher–child interactions and preschool children's school readiness skills. *Early Childhood Research Quarterly*, Q3 2016, 36, 561-571. doi:10.1016/j.ecresq.2015.09.005

Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, 25, 166-176. doi:10.1016/j.ecresq.2009.10.004

Magnusson & Lin, 2016: http://dcf.wisconsin.gov/youngstar/pdf/FINAL-Part-2-Validation-Study-Full-Report-2016.pdf

Soliday Hong, S. L., Howes, C., Marcella, J., Zucker, E., & Huang, Y. (2015). Quality Rating and Improvement Systems: Validation of a local implementation in LA County and children's school-readiness. *Early Childhood Research Quarterly*, 30, Part B, 227-240. http://dx.doi.org/10.1016/j.ecresq.2014.05.001

Tout, K., Cleveland, J., Li, W., Starr, R., Soli, M., & Bultinck, E. (2016). The Parent Aware Evaluation: Initial Validation Report. Minneapolis, MN: Child Trends.

Appendices

Appendix A. Oregon QRIS Domains and Standards

Domain: Learning and Development (10/12 needed to pass at any star level)

LD1: The program is guided by a written statement of philosophy.
LD2: The program uses a curriculum that supports all children's learning and development.
LD3: The program provides an appropriate indoor environment that supports children's learning and development and is accessible to all children enrolled in the program.
LD4: The program provides appropriate indoor furnishings that support children's learning and development.
LD5: The program provides appropriate and well-maintained outdoor gross motor area with equipment.
LD6: The program uses materials that support children's learning and development.
LD7: The program uses planned curriculum activities that support children's learning and development.
LD8 The program uses daily routines that support children's learning and development.
LD9: The program uses information from screening and assessment to measure children's learning and development in order to make referrals and do program planning.
LD10: Group size, child-staff ratios, and staffing patterns are appropriate for the children's age and positively affect children's emotional development, cognitive development, safety, and health.
LD11: The program facilitates and supports appropriate adult-child interactions in the areas of social and emotional support, organization and management of children's behavior, and instructional support.
LD12: The program facilitates and supports children's positive social and emotional development

Domain: Health and Safety (5/6 needed to pass at any star level)

HS1: Children are provided instruction and support to independently manage health and hygiene practices.
HS2: Children are provided instruction and support on safety rules and expectations.
HS3: Healthy eating habits are supported and encouraged.
HS4: Healthy fitness habits are supported and encouraged.
HS5: Program personnel collaborate with health and related service professionals to address the individual health needs of children as applicable.
HS6: Program uses screen time appropriately. Screen time includes all electronic media such as television, video/DVD, electronic games, computers, tablets, smart phones, or any other screened electronic devices.

Domain: Personnel Qualifications (5/5 needed to pass at any star level)

PQ1: The program's leader is presently qualified through education, training, and experience.
PQ2: The program's head teacher(s) and teacher(s) are presently qualified to serve in their positions through education, training, and experience. *(n/a for Certified Family and Registered Family programs)*
PQ3: The program's assistants and/or aides are presently qualified to serve in their positions through education, training, and experience.
PQ4: Program personnel continue to advance their knowledge and skills through participation in training and/or college course credits annually that are part of a professional development plan that will lead to advancement up to Step 10 on the Oregon Registry.

PQ5: Program personnel are trained in ethics, professional responsibility, and maintaining confidentiality.

Domain: Family Partnerships (3/4 needed to pass at any star level)

FP1: The program uses family input and feedback to guide program planning and policy decisions.
FP2: The program meets the individual needs of children through mutually respectful, two-way communication with families.
FP3: Families are encouraged to be regular and frequent participants in the program.
FP4: The program provides support and information to assist families in meeting their child's needs and goals.

Domain: Administration and Business Practices (5/6 needed to pass at any star level):

AB1: The program follows sound business practices, policies, and procedures that support financial sustainability.
AB2: In programs were there are multiple employees, the program assures a professional working climate.
AB3: In programs were there are multiple employees, personnel are evaluated on their performance.
AB4: In programs were there are multiple employees, the program promotes positive working relationships and professionalism.
AB5: A comprehensive program evaluation process is developed and performed on an annual basis. The evaluation examines the program's policies and procedures, care and education environment, curriculum, and administration and business practices.
AB6: In programs were there are multiple employees, benefits are offered to encourage retention.

For a full list of specific indicators for each standard, for Centers and Family programs see: http://triwou.org/projects/qris/earlylearning

Appendix B. Crosswalk of QRIS Standards with Head Start and NAEYC

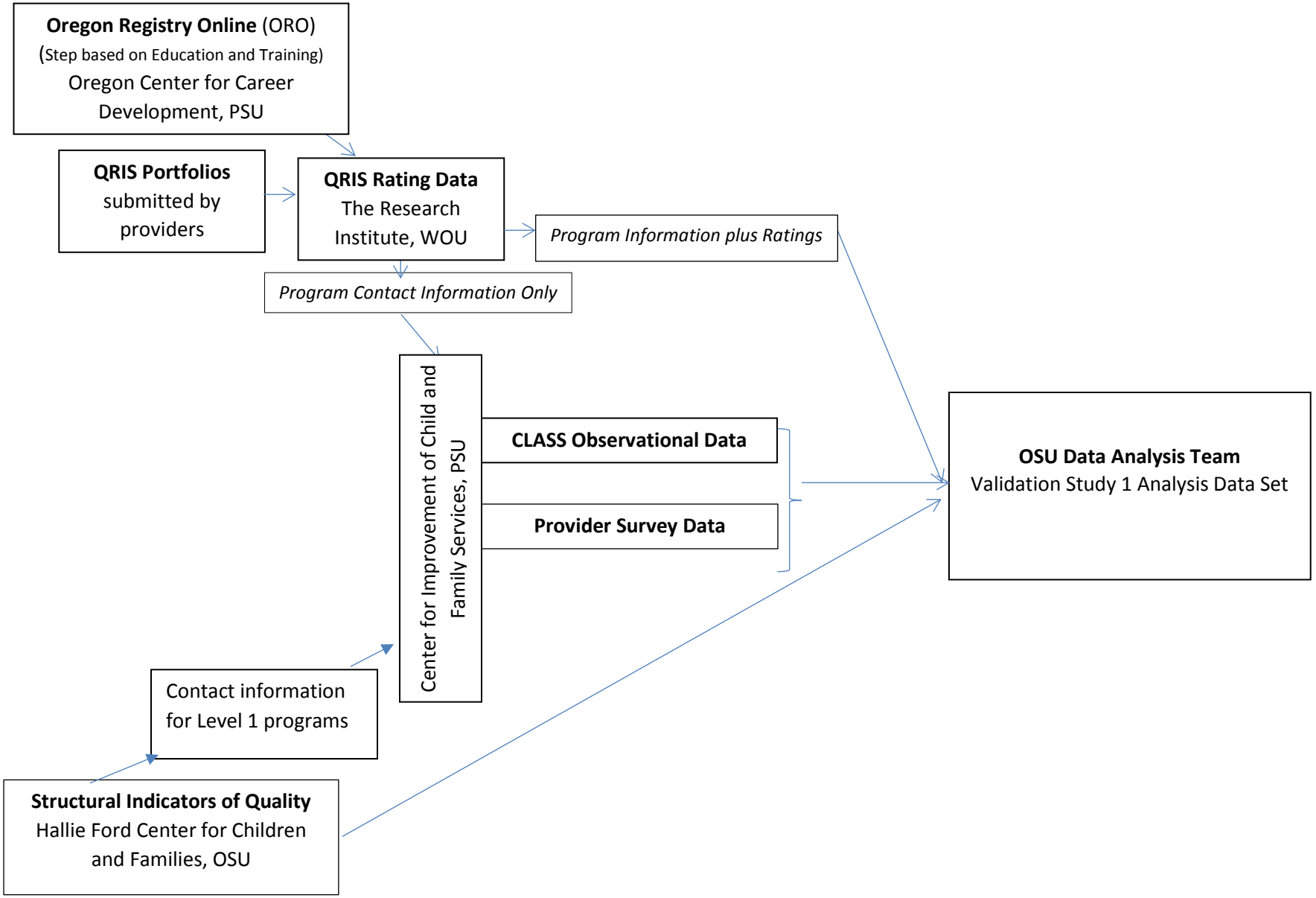
Cross-Walk of Oregon's QRIS and Head Start and Accrediting Bodies

QRIS standards programs ***must address*** in order to receive a five-star QRIS rating. Standards not noted in the Table are considered to be automatically met by Head Start or Accreditation.

	Head Start	National Association for the Education of Young Children	Association of Christian Schools International	National Association for Family Child Care	Oregon Program of Quality
Learning and Development					
LD1				3-star, 4-star, & 5-star	4-star & 5-star
LD2				3-star	
LD3			4-star & 5-star	4-star	
LD4		5-star	4-star & 5-star	4-star & 5-star	4-star & 5-star
LD5					
LD6				5-star	
LD7			5-star		4-star & 5-star
LD8					
LD9		3-star	3-star	3-star, 4-star, & 5-star	3-star
LD10			4-star & 5-star	4-star	3-star, 4-star, & 5-star
LD11	5-star (grantee)	5-star	5-star	5-star	3-star, 4-star, & 5-star
LD12	5-star (grantee)	5-star	5-star	5-star	3-star, 4-star, & 5-star
Health and Safety					
HS1			3-star & 5-star	5-star	
HS2				3-star	
HS3		4-star		4-star & 5-star	
HS4					3-star & 4-star
HS5					
HS6	3-star & 5-star grantee	3-star & 5-star			3-star & 5-star
Personnel Qualifications					
PQ1	5-star (site)	3-star, 4-star, & 5-star	5-star	3-star, 4-star, & 5-star	3-star, 4-star, & 5-star
PQ2	5-star (site)	3-star, 4-star, & 5-star	5-star		3-star, 4-star, & 5-star
PQ3	5-star (site)	3-star, 4-star, & 5-star	5-star	3-star, 4-star, & 5-star	3-star, 4-star, & 5-star

PQ4	5-star (site)	3-star, 4-star, & 5-star	5-star	3-star, 4-star, & 5-star	3-star, 4-star, & 5-star
PQ5			4-star	3-star	
Family Partnership					
FP1				5-star	5-star
FP2					
FP3					
FP4					
Administration and Business Practices					
AB1				3-star & 4-star	
AB2				4-star	
AB3	4-star grantee		5-star	4-star & 5-star	
AB4			5-star	3-star, 4-star, & 5-star	3-star
AB5		5-star		4-star & 5-star	5-star
AB6	5-star (grantee)		5-star	5-star	5-star

Appendix C. Oregon's QRIS Validation Study Data Sources



Appendix D. Behavioral Markers for Combined CLASS Instrument (PreK and Toddler)

CLASS Toddler/PreK Combined Behavioral Markers

Positive Climate

- Relationships
  - Physical proximity
  - Shared activities
  - Peer assistance and peer connections
  - Matched affect
  - Social conversation
  - Reciprocal interaction
- Positive affect
  - Smiling
  - Laughter
  - Enthusiasm
- Positive communication
  - Verbal affection
  - Physical affection
  - Positive expectations
- Respect
  - Eye contact
  - Warm, calm voice
  - Respectful language & communication
  - Cooperation and/or sharing
  - Body orientation
  - Communicate before physically moving child

Negative Climate

- Negative affect
  - Irritability
  - Anger
  - Harsh voice
- Punitive control
  - Yelling
  - Threats
  - Physical control, actions, and punishment
  - Harsh punishment
- Teacher negativity
  - Sarcastic voice/statement
  - Teasing
  - Humiliation
- Child negativity
  - Peer disputes
  - Escalating frustration
  - Disconnected or escalating negativity
- Severe negativity
  - Victimization
  - Bullying

Teacher Sensitivity

Awareness

- Anticipates problems and plans accordingly
- Is attentive throughout the classroom
- Notices lack of understanding, difficulties, and/or children who are upset

Responsiveness

- Acknowledges and accepts emotions
- Provides comfort and assistance
- Provides individualized support and comfort
- Responds to children's bids for attention

Addresses problems

- Helps in an effective and timely manner
- Helps to genuinely resolve problems

Student comfort

- Seeks support and guidance
- Freely approaches and participates
- Takes risks

Regard for Student Perspectives

Flexibility and student focus

- Shows flexibility and/or "goes with the flow"
- Incorporates students' ideas
- Follows students' leads
- Adjusts pacing for individual children

Support for autonomy and leadership

- Allows choice
- Allows students to lead lessons
- Gives students responsibility
- Support of self-care
- Materials accessible
- Peer perspective taking

Student expression

- Encourages student talk
- Elicits ideas, perspectives, and/or expression

Restriction of movement

- Allows movement and talking
- Is not rigid

Behavior Management

- Clear Behavior Expectations
  - Clear expectations
  - Consistency
  - Clarity of rules
  - Children demonstrate awareness of expectations
- Proactive
  - Anticipates problem behavior or escalation
  - Low reactivity
  - Actively monitors children's behavior
- Redirection of misbehavior
  - Effective reduction of misbehavior
  - Uses subtle cues to redirect
  - Efficient redirection
  - Specificity in redirection and/or directions
- Supporting positive behavior
  - Attention to the positive
  - Reinforcement of positive behavior
  - Positive phrasing of desired behavior
- Student behavior and problem behavior
  - Frequent compliance
  - Little aggression and defiance
  - Minimal wandering
  - Minimal waiting
  - Lack of disruptive or potentially dangerous behavior

Productivity

- Maximizing learning time
  - Provision of activities
  - Choice when finished
  - Few disruptions
  - Effective completion of managerial tasks
  - Pacing
- Routines
  - Students know what to do
  - Clear instructions
  - Little wandering
- Transitions
  - Brief
  - Explicit follow-through
  - Learning opportunities within
- Preparation
  - Materials ready and accessible
  - Knows lessons

Instructional Learning Formats

Effective facilitation
- Teacher involvement
- Effective questioning
- Expanding children's involvement

Variety of modalities and materials
- Range of auditory, visual, and movement opportunities
- Interesting and creative materials
- Hands-on opportunities

Student interest
- Active participation
- Listening
- Focused attention

Clarity of learning objectives
- Advanced organizers
- Summaries
- Reorientation statements

Concept Development

Analysis and reasoning
- Why and/or how questions
- Problem solving
- Prediction/experimentation
- Classification/comparison
- Evaluation

Creating
- Brainstorming
- Planning
- Producing

Integration
- Connects concepts
- Integrates with previous knowledge

Connections to the real world
- Real-world applications
- Related to students' lives

Facilitation of Learning and Development

Active facilitation
- Teacher provides opportunities for exploration and learning
- Teacher guides exploration
- Teacher is involved in children's activities to support learning and development

Expansion of cognition
- Teacher provides and embeds information
- Teacher relates information to children's lives and experiences
- Teacher integrates concepts across activities and tasks
- Teacher encourages thinking skills

Children's active engagement
- Manipulation of materials
- Physical involvement
- Verbal involvement

Quality Feedback

Scaffolding
- Hints
- Verbal or physical assistance

Feedback loops
- Back-and-forth exchanges
- Persistence by teacher
- Follow-up questions

Prompting thought processes
- Asks students to explain thinking
- Queries response and actions

Providing information
- Expansion and elaboration
- Clarification of concepts and tasks
- Specific feedback

Encouragement and affirmation
- Recognition of effort or accomplishment
- Reinforcement
- Student persistence
- Individualized feedback

Language Modeling

Frequent conversation and supporting language use
- Back-and-forth exchanges
- Contingent responding
- Peer conversations

Open-ended questions
- Questions require more than one-word answer
- Students respond

Repetition and extension
- Repeats
- Extends/elaborates

Self- and parallel-talk
- Maps own actions with language
- Maps student action with language

Advanced language
- Variety of words and/or descriptive vocabulary
- Labeling
- Connected to familiar words and/or ideas
- Teachers give children words to say

Appendix E. Supplemental Results for the Oregon QRIS Validation Study

Appendix E contains additional results to supplement those presented in the full report. Results are organized by the 5 Research Questions.

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

Table 5a. Frequencies: QRIS ratings in Registered Family programs (N=45)

			# (%) Rated at each star-level			
Domains & Standards	**Total**	**Missing**	**2**	**3**	**4**	**5**
Domain: Learning & Development						
Learning Development 1	45	0	13(28.9)	24(53.3)	6(13.3)	2(4.4)
Learning Development 2	45	0	7(15.6)	27(60.0)	6(13.3)	5(11.1)
Learning Development 3	45	0	4(8.9)	31(68.9)	7(15.6)	3(6.7)
Learning Development 4	45	0	7(15.6)	27(60.0)	8(17.8)	3(6.7)
Learning Development 5	45	0	5(11.1)	26(57.8)	7(15.6)	7(15.6)
Learning Development 6	45	0	6(13.3)	29(64.4)	7(15.6)	3(6.7)
Learning Development 7	45	0	14(31.1)	21(46.7)	7(15.6)	3(6.7)
Learning Development 8	45	0	6(13.3)	30(66.7)	6(13.3)	3(6.7)
Learning Development 9	45	0	23(51.1)	18(40.0)	2(4.4)	2(4.4)
Learning Development 10	45	0	2(4.4)	32(71.1)	9(20.0)	2(4.4)
Learning Development 11	45	0	19(42.2)	18(40.0)	6(13.3)	2(4.4)
Learning Development 12	45	0	9(20.0)	26(57.8)	8(17.8)	2(4.4)
Domain: Health & Safety						
Health & Safety 1	45	0	15(33.3)	19(42.2)	9(20.0)	2(4.4)
Health & Safety 2	45	0	1(2.2)	34(75.6)	8(17.8)	2(4.4)
Health & Safety 3	45	0	12(26.7)	22(48.9)	8(17.8)	3(6.7)
Health & Safety 4	45	0	4(8.9)	29(64.4)	8(17.8)	4(8.9)
Health & Safety 5	45	0	3(6.7)	32(71.1)	6(13.3)	4(8.9)
Health & Safety 6	45	0	14(31.1)	21(46.7)	7(15.6)	3(6.7)
Domain: Personnel Qualifications						
Personnel Qualifications 1	45	0	12(26.7)	24(53.3)	4(8.9)	5(11.1)
Personnel Qualifications 2	45	34(75.6)	0(0.0)	8(17.8)	2(4.4)	1(2.2)
Personnel Qualifications 3	45	32(71.1)	0(0.0)	10(22.2)	2(4.4)	1(2.2)
Personnel Qualifications 4	45	0	4(8.9)	27(60.0)	8(17.8)	6(13.3)
Personnel Qualifications 5	45	0	1(2.2)	34(75.6)	7(15.6)	3(6.7)
Domain: Family Partnerships						

Family Partnerships 1	45	0	3(6.7)	32(71.1)	9(20.0)	1(2.2)
Family Partnerships 2	45	0	3(6.7)	33(73.3)	6(13.3)	3(6.7)
Family Partnerships 3	45	0	0(0.0)	31(68.9)	10(22.2)	4(8.9)
Family Partnerships 4	45	0	3(6.7)	31(68.9)	8(17.8)	3(6.7)
Domain: Administration & Business Practice						
Administration & Business Practice 1	45	0	1(2.2)	34(75.6)	6(13.3)	4(8.9)
Administration & Business Practice 2	45	32(71.1)	0(0.0)	9(20.0)	2(4.4)	2(4.4)
Administration & Business Practice 3	45	32(71.1)	0(0.0)	9(20.0)	2(4.4)	2(4.4)
Administration & Business Practice 4	45	32(71.1)	0(0.0)	9(20.0)	2(4.4)	2(4.4)
Administration & Business Practice 5	45	0	16(35.6)	19(42.2)	8(17.8)	2(4.4)
Administration & Business Practice 6	45	31(68.9)	0(0.0)	9(20.0)	3(6.67)	2(4.44)

[1]Minimum score for all Standards is 2 and maximum is 5.

Table 5b. Frequencies: QRIS ratings in Certified Centers (N=120)

			# (%) Rated at each star-level			
Domain & Standards	**Total**	**Missing**	**2**	**3**	**4**	**5**
Domain: Learning & Development						
Learning Development 1	120	0	19(15.8)	31(25.8)	14(11.7)	56(46.7)
Learning Development 2	120	0	11(9.2)	31(25.8)	9(7.5)	69(57.5)
Learning Development 3	120	0	12(10.0)	38(31.7)	15(12.5)	55(45.8)
Learning Development 4	120	0	15(12.5)	37(30.8)	11(9.2)	57(47.5)
Learning Development 5	120	0	7(5.8)	34(28.3)	11(9.2)	68(56.7)
Learning Development 6	120	0	15(12.5)	38(31.7)	10(8.3)	57(47.5)
Learning Development 7	120	0	23(19.2)	29(24.2)	12(10.0)	56(46.7)
Learning Development 8	120	0	11(9.2)	36(30.0)	12(10.0)	61(50.8)
Learning Development 9	120	0	50(41.7)	22(18.3)	4(3.3)	44(36.7)
Learning Development 10	120	0	10(8.3)	38(31.7)	12(10.0)	60(50.0)
Learning Development 11	120	0	26(21.7)	34(28.3)	40(33.3)	20(16.7)
Learning Development 12	120	0	10(8.3)	44(36.7)	23(19.2)	43(35.8)
Domain: Health & Safety						
Health & Safety 1	120	0	21(17.5)	33(27.5)	10(8.3)	56(46.7)
Health & Safety 2	120	0	4(3.3)	44(36.7)	9(7.5)	63(52.5)
Health & Safety 3	120	0	19(15.8)	38(31.7)	8(6.7)	55(45.8)
Health & Safety 4	120	0	5(4.2)	42(35.0)	11(9.2)	62(51.7)
Health & Safety 5	120	0	18(15.0)	32(26.7)	11(9.2)	59(49.2)
Health & Safety 6	120	0	66(55.0)	17(14.2)	12(10.0)	25(20.8)
Domain: Personnel Qualifications						

Personnel Qualifications 1	120	0	19(15.8)	32(26.7)	15(12.5)	54(45.0)
Personnel Qualifications 2	120	1(0.8)	22(18.3)	26(21.7)	14(11.7)	57(47.5)
Personnel Qualifications 3	120	39(32.5)	13(10.8)	18(15.0)	10(8.3)	40(33.3)
Personnel Qualifications 4	120	0	15(12.5)	28(23.3)	17(14.2)	60(50.0)
Personnel Qualifications 5	120	0	12(10.0)	45(37.5)	7(5.8)	56(46.7)
Domain: Family Partnerships						
Family Partnerships 1	120	0	12(10.0)	39(32.5)	17(14.2)	52(43.3)
Family Partnerships 2	120	0	7(5.8)	48(40.0)	13(10.8)	52(43.3)
Family Partnerships 3	120	0	2(1.7)	36(30.0)	20(16.7)	62(51.7)
Family Partnerships 4	120	0	7(5.8)	41(34.2)	9(7.5)	63(52.5)
Domain: Administration & Business Practice						
Administration & Business Practice 1	120	2(1.67)	3(2.50)	42(35.0)	7(5.83)	66(55.0)
Administration & Business Practice 2	120	1(0.8)	10(8.33)	36(30.0)	9(7.50)	64(53.3)
Administration & Business Practice 3	120	1(0.8)	19(15.8)	38(31.7)	12(10.0)	50(41.7)
Administration & Business Practice 4	120	1(0.8)	6(5.0)	36(30.0)	10(8.3)	67(55.8)
Administration & Business Practice 5	120	0	19(15.8)	36(30.0)	11(9.2)	54(45.0)
Administration & Business Practice 6	120	1(0.8)	4(3.3)	35(29.2)	12(10.0)	68(56.7)

[1]Minimum score for all Standards is 2 and maximum is 5.

Table 5c. Frequencies: QRIS ratings in Certified Family programs (N=81)

			# (%) Rated at each star-level			
Domain & Standards	**Total**	**Missing**	**2**	**3**	**4**	**5**
Domain: Learning & Development						
Learning Development 1	81	0	12(14.8)	32(39.5)	14(17.3)	23(28.4)
Learning Development 2	81	0	10(12.3)	27(33.3)	11(13.6)	33(40.7)
Learning Development 3	81	0	5(6.2)	37(45.7)	14(17.3)	25(30.9)
Learning Development 4	81	0	6(7.4)	39(48.1)	10(12.3)	26(32.1)
Learning Development 5	81	0	4(4.9)	31(38.3)	12(14.8)	34(42.0)
Learning Development 6	81	0	5(6.2)	37(45.7)	14(17.3)	25(30.9)
Learning Development 7	81	0	9(11.1)	30(37.0)	12(14.8)	30(37.0)
Learning Development 8	81	0	5(6.2)	32(39.5)	14(17.3)	30(37.0)
Learning Development 9	81	0	31(38.3)	28(34.6)	6(7.4)	16(19.8)
Learning Development 10	81	0	3(3.7)	38(46.9)	13(16.0)	27(33.3)
Learning Development 11	81	0	24(29.6)	26(32.1)	12(14.8)	19(23.5)
Learning Development 12	81	0	10(12.3)	37(45.7)	12(14.8)	22(27.2)
Domain: Health & Safety						
Health & Safety 1	81	0	14(17.3)	28(34.6)	13(16.0)	26(32.1)

Health & Safety 2	81	0	5(6.2)	39(48.1)	8(9.9)	29(35.8)
Health & Safety 3	81	0	7(8.6)	40(49.4)	11(13.6)	23(28.4)
Health & Safety 4	81	0	1(1.2)	40(49.4)	8(9.9)	32(39.5)
Health & Safety 5	81	0	11(13.6)	33(40.7)	9(11.1)	28(34.6)
Health & Safety 6	81	0	23(28.4)	26(32.1)	7(8.6)	25(30.9)
Domain: Personnel Qualifications						
Personnel Qualifications 1	81	0	9(11.1)	29(35.8)	14(17.3)	29(35.8)
Personnel Qualifications 2	81	51(63.0)	1(1.2)	12(14.8)	4(4.9)	13(16.1)
Personnel Qualifications 3	81	22(27.2)	5(6.2)	18(22.2)	10(12.4)	26(32.1)
Personnel Qualifications 4	81	0	8(9.9)	27(33.3)	11(13.6)	35(43.2)
Personnel Qualifications 5	81	0	8(9.9)	35(43.2)	9(11.1)	29(35.8)
Domain: Family Partnerships						
Family Partnerships 1	81	0	7(8.6)	36(44.4)	23(28.4)	15(18.5)
Family Partnerships 2	81	0	3(3.7)	45(55.6)	10(12.3)	23(28.4)
Family Partnerships 3	81	0	0(0.0)	35(43.2)	15(18.5)	31(38.3)
Family Partnerships 4	81	0	2(2.5)	38(46.9)	12(14.8)	29(35.8)
Domain: Administration & Business Practice						
Administration & Business Practice 1	81	0	5(6.2)	39(48.1)	9(11.1)	28(34.6)
Administration & Business Practice 2	81	26(32.1)	6(7.4)	18(22.2)	9(11.1)	22(27.2)
Administration & Business Practice 3	81	27(33.3)	5(6.2)	20(24.7)	12(14.8)	17(21.0)
Administration & Business Practice 4	81	28(34.6)	5(6.17)	17(21.0)	10(12.4)	21(25.9)
Administration & Business Practice 5	81	1(1.23)	17(20.99)	31(38.3)	9(11.1)	23(28.4)
Administration & Business Practice 6	81	26(32.1)	1(1.2)	20(24.7)	12(14.8)	22(27.2)

[1]Minimum score for all Standards is 2 and maximum is 5.

Table 6a. Descriptive Statistics: QRIS Ratings (all programs)

Standards	N	Mode	Median	Mean	SD
Learning Development 1	246	3	3	3.62	1.12
Learning Development 2	246	5	4	3.86	1.11
Learning Development 3	246	3	3	3.74	1.02
Learning Development 4	246	3	3	3.70	1.07
Learning Development 5	246	5	4	3.94	1.04
Learning Development 6	246	3	3	3.71	1.06
Learning Development 7	246	5	3	3.66	1.15
Learning Development 8	246	3	4	3.80	1.05
Learning Development 9	246	2	3	3.13	1.21

Learning Development 10	246	3	3.5	3.80	1.01
Learning Development 11	246	3	3	3.29	1.05
Learning Development 12	246	3	3	3.60	1.01
Average Learning Development	**246**	**n/a**	**3.6**	**3.66**	**0.94**
Health & Safety 1	246	5	3	3.61	1.15
Health & Safety 2	246	3	3	3.83	1.00
Health & Safety 3	246	3	3	3.61	1.10
Health & Safety 4	246	3	4	3.87	1.00
Health & Safety 5	246	3	3	3.72	1.10
Health & Safety 6	246	2	3	3.12	1.17
Average Health & Safety	**246**	**n/a**	**3.2**	**3.62**	**0.91**
Personnel Qualifications 1	246	5	3	3.69	1.12
Personnel Qualifications 2	160	5	4	3.87	1.14
Personnel Qualifications 3	153	5	4	3.90	1.10
Personnel Qualifications 4	246	5	4	3.86	1.08
Personnel Qualifications 5	246	3	3	3.72	1.04
Average Personnel Qualifications	**246**	**n/a**	**3.7**	**3.76**	**0.94**
Family Partnerships 1	246	3	3	3.66	0.98
Family Partnerships 2	246	3	3	3.70	0.98
Family Partnerships 3	246	3	4	3.96	0.92
Family Partnerships 4	246	3	4	3.84	1.00
Average Family Partnerships	**246**	**n/a**	**3.5**	**3.79**	**0.90**
Administration & Business Practice 1	244	3	3	3.86	1.00
Administration & Business Practice 2	187	5	4	3.96	1.07
Administration & Business Practice 3	186	5	4	3.75	1.09
Administration & Business Practice 4	185	5	4	4.03	1.03
Administration & Business Practice 5	245	3	3	3.54	1.15
Administration & Business Practice 6	188	5	4	4.10	0.97
Administration & Business Practice	**246**	**5**	**3.5**	**3.74**	**0.97**

[1]Minimum score for all Standards is 2 and maximum is 5.

Table 6b. Descriptive Statistics: CLASS scores by program type

CLASS Scores Average across all 3 Instruments	N	Minimum	Maximum	Median	Mean	SD
Certified Centers						
Emotional Support[1]	126	2.85	6.83	5.39	5.37	0.67
Organized Classrooms[2]	113	2.44	6.00	4.67	4.60	0.70
Instructional Support	126	1.37	5.61	2.63	2.66	0.71
Total	126	2.40	6.27	4.24	4.30	0.62
Certified Family						
Emotional Support[1]	83	3.04	6.75	5.58	5.46	0.76
Organized Classrooms[2]	66	2.78	6.78	4.78	4.86	0.86
Instructional Support	83	1.33	5.25	2.75	2.82	0.83
Total	83	2.76	6.26	4.36	4.39	0.71
Registered Family						
Emotional Support[1]	47	3.75	6.83	5.67	5.46	0.77
Organized Classrooms[2]	37	2.11	6.00	4.67	4.55	0.90
Instructional Support	47	1.00	4.67	2.56	2.64	0.84
Total	47	2.29	6.03	4.22	4.28	0.75

[1]Emotional Support for Toddler Measure Includes Behavioral Guidance

[2]Toddler Measure does not Include Organization Support. Behavioral Guidance included in Emotional Support

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

Table 7a. Correlation within QRIS standards: Learning Development

		Learning Development											
		Written philosophy	Uses curriculum	Appropriate Indoor Environment	Indoor Furnishings	Outdoor Environment	Materials Support Development	Planned Curriculum Activities	Routines Support I&D	Screening	Group Size/Ratio	Adult Child Interactions	Supports Soc. Emotional Dev.
Learning Development	Written Philosophy	3											
	Uses Curriculum	.731***	1										
	Appropriate Indoor Environment	.813***	.726***	1									
	Indoor Furnishings	.713***	.712***	.852***	1								
	Outdoor Environment	.705***	.841***	.807***	.807***	1							
	Materials Support Dev	.766***	.711***	.872***	.862***	.791***	1						
	Planned Curriculum Activities	.740***	.749***	.846***	.781***	.794***	.817***	1					
	Routines	.747***	.781***	.841***	.781***	.795***	.836***	.839***	1				
	Screening	.670***	.553***	.667***	.654***	.536***	.691***	.611***	.626***	1			
	Group Size/Ratio	.769***	.762***	.799***	.763***	.761***	.785***	.778***	.833***	.615***	1		
	Adult Child Interactions	.700***	.555***	.718***	.670***	.577***	.724***	.695***	.680***	.656***	.650***	1	
	Supports Soc. Em. Dev.	.746***	.702***	.825***	.778***	.718***	.813***	.774***	.801***	.715***	.769***	.815***	1

***Correlation is significant at the $p < .0001$ level (2-tailed)

Table 7b. Correlations within QRIS Standards: Health & Safety

		Health & Safety					
		Instruction on Health/ Hygiene	Instruction on Safety	Support Eating Habits	Support Fitness Habits	Collaborate with H&S Profs	Screen Time
Health & Safety	Instruction on Health/Hygiene	1					
	Instruction on Safety	.664***	1				
	Support Eating Habits	.694***	.742***	1			
	Support Fitness Habits	.736***	.780***	.774***	1		
	Collaborate with H&S Profs	.703***	.792***	.763***	.753***	1	
	Screen Time	.432***	.426***	.465***	.459***	.485***	1

***Correlation is significant at the $p < .0001$ level (2-tailed)

Table 7c. Correlations within QRIS standards: Personnel Qualifications

		Personnel Qualifications				
		Lead Qualified	Teachers Qualified[1]	Aide is Qualified[2]	Training to Get to Step 10	Personnel Trained in Ethics.
Personnel Qualifications	Lead Qualified	1				
	Teachers Qualified[1]	.636***	1			
	Aide is Qualified[2]	.706***	.621***	1		
	Training to Get to Step 10	.660***	.695***	.598***	1	
	Personnel Trained in Ethics.	.653***	.741***	.609***	.685***	1

***Correlation is significant at the $p < .0001$ level (2-tailed)

[1] n= 160, [2] n=153

Table 7d. Correlations within QRIS standards: Family Partnership

		Family Partnerships			
		Use Family Input	2-way Communication	Participation Encouraged	Provides Info to Assist Fam in Meeting Child's Needs/Goals
Family Partnerships	Use Family Input	1			
	2-way Communication	.819***	1		
	Participation Encouraged	.767***	.811***	1	
	Provides Info to Assist fam in Meeting Child's Needs & Goals	.776***	.821***	.870***	1

***Correlation is significant at the $p < .0001$ level (2-tailed)

Table 7e. Correlations within QRIS standards: Administration and Business Practice

		Administration and Business Practice					
		Sustainable Practices	Prof Working Climate	Personnel Evaluations	Promotes Pos. Working Relationships	Annual Evaluations	Benefits
Administration and Business Practice	Sustainable Practices	1[1]					
	Prof Working Climate	.827***	1[2]				
	Personnel Evaluations[3]	.692***	.748***[3]	1[3]			
	Promotes Pos. Working Relationships[4]	.832***[4]	.839***[5]	.693***[5]	1[5]		
	Annual Evaluations[5]	.774***	.734***[2]	.650***[3]	.711***[5]	1	
	Benefits[6]	.798***[3]	.762***[5]	.686***[5]	.820***[5]	.678***[2]	1

***Correlation is significant at the $p < .001$ level (2-tailed)

[1] n=244; [2] n=187; [3] n= 186; [4] n=183 [5]n=185; [6] n=245; [7] n=188

Table 7f. Alpha Coefficients (Internal Consistency)

	Coefficient	Items	N
Overall (All Standards)[1,2]	0.99	33	115
Learning Development	0.97	12	246
Health & Safety	0.91	6	246
Personnel Qualifications[2]	0.92	5	120
Family Partnerships	0.94	4	246
Administration & Business Practice[2]	0.93	5	185

[1]Uses listwise deletion

[2]When responses were coded as Not Applicable they were recoded as missing.

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

Table 9a. Analysis of Variance (ANOVA) tests of differences in CLASS scores by QRIS rating (including Level 1 programs)

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS Instrument	**CLASS domains**			**CLASS domains**			**CLASS domains**			**CLASS domains**		
	ES	**IS**	**OC**	**ES**	**IS**	**OC**	**ES**	**IS**	**OC**	**ES**	**IS**	**OC**
Overall												
each rating	**2.00+**	**3.99****	**3.40***	1.71	1.00	1.67	**2.02+**	1.24	**2.54***	1.24	1.50	**2.26+**
1-2 vs 3-5	**5.22***	**14.66****	**11.54****	1.40	**3.67+**	0.52	1.75	**3.45+**	**7.26****	1.58	**5.23***	**6.65***
N	304	304	259	149	149	136	92	92	74	63	63	49
PreK												
each rating	**2.23+**	**4.80****	1.72	1.38	1.93	0.96	0.94	0.90	1.33	0.34	**3.33+**	1.53
1-2 vs 3-5	**6.85***	**14.93****	**5.96***	2.46	**5.07***	0.10	**3.75+**	**3.21+**	**4.79***	0.05	**4.15+**	**3.80+**
N	195	195	195	131	131	131	37	37	37	27	27	27
Toddler												
each rating	0.87	**2.31+**	n/a	1.93	**2.01+**	n/a	0.32	0.68	n/a	^	^	n/a
1-2 vs 3-5	0.19	**5.31***		0.96	**5.26***		0.00	0.00				
N	126	126		84	84		26	26				
Combined												
each rating	**2.12+**	1.66	**2.23+**	^	^	^	1.66	0.70	1.10	1.25	1.21	1.00
1-2 vs 3-5	**6.23***	**5.93***	**8.17****				1.04	**2.88+**	**3.69+**	**3.48+**	2.53	2.29
N	80	80	80				41	41	41	25	25	25

Note. For each CLASS instrument two different ANOVAs were conducted: one with each QRIS rating kept as a separate group, and one with programs rated 1-2 grouped together and compared against programs rated 3-5 as a second group. Estimates in the table are the F values from the ANOVA tests.

+ Nearing significance at the $p < .10$ level (2-tailed)

* Statistically significant at the $p < .05$ level (2-tailed)

** Statistically significant at the $p < .01$ level (2-tailed)

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Table 9b. Analysis of Variance (ANOVA) tests of differences in CLASS scores by QRIS rating (does not include Level 1 programs)

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS Instrument	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
All (average)	0.90	**2.87***	**4.55****	0.74	0.87	1.43	1.48	1.17	1.88	1.71	1.01	1.12
2 vs 3-5	1.76	**8.50****	**11.71****	0.09	2.48	**3.79^{+}**	0.43	2.01	**3.29^{+}**	2.29	**2.68+**	2.14
N	246	246	206	120	120	107	81	81	64	45	45	35
PreK	1.50	**2.98***	**2.32^{+}**	0.48	1.87	0.72	1.07	0.74	1.10	^	^	^
2 vs 3-5	**3.35^{+}**	**7.10****	**6.32****	0.52	**3.14^{+}**	1.42	**3.07^{+}**	1.81	2.79			
N	156	156	156	103	103	103	33	33	33			
Toddler	0.85	**3.16***	n/a	0.91	2.02	n/a	0.37	0.83	n/a	^	^	n/a
2 vs 3-5	0.00	**5.21***		0.02	**3.58^{+}**		0.13	0.14				
N	104	104		70	70		23	23				
Combined	1.02	1.66	**2.19^{+}**	^	^	^	0.32	0.41	0.60	^	^	^
2 vs 3-5	1.99	**4.27***	**5.58***				0.30	1.82	0.96			
N	63	63	63				34	34	34			

Note. For each CLASS instrument two different ANOVAs were conducted: one with each QRIS rating kept as a separate group, and one with programs rated 1-2 grouped together and compared against programs rated 3-5 as a second group. Estimates in the table are the F values from the ANOVA tests.

$^{+}$ Nearing significance at the $p < .10$ level (2-tailed)

* Statistically significant at the $p < .05$ level (2-tailed)

** Statistically significant at the $p < .01$ level (2-tailed)

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Table 9c. Effect Size (Cohen's d) for ANOVAs comparing programs rated 1-2 and those rated 3-star or higher

	CLASS Domain	Cohen's d
PreK	Emotional Support	.37
	Organized Classrooms	.35
	Instructional Support	.56
	Total	.49
Toddler	Emotional Support & Behavior	.08 (ns)
	Instructional Support	.41
	Total	.21 (ns)
Combined	Emotional Support	.57
	Organized Classrooms	.65
	Instructional Support	.55
	Total	.68
Total	Emotional Support	.26
	Organized Classrooms	.42
	Instructional Support	.44
	Total	.37

Analyses included all three program types: Certified Centers, Certified Family, Registered Family.

Table 11a. For programs with each QRIS rating how many had low, medium, and high CLASS scores? (PreK)

		QRIS Ratings			
CLASS		2	3	4	5
Emotional Support					
	Low	16(27%)	**9(23%)**	**2(14%)**	**11(26%)**
	Medium	32(53%)	17(44%)	6(43%)	23(53%)
	High	**12(20%)**	13(33%)	6(43%)	9(21%)
Organized Classrooms					
	Low	13(22%)	**4(10%)**	0(0%)	**7(16%)**
	Medium	33(55%)	20(51%)	8(57%)	12(28%)
	High	**14(23%)**	15(38%)	6(43%)	24(56%)
Instructional Support					
	Low	16(27%)	**11(28%)**	**1(7%)**	**3(7%)**
	Medium	34(57%)	15(38%)	6(43%)	23(53%)
	High	**10(17%)**	13(33%)	7(50%)	17(40%)

Notes. The cut-off points used to create high, medium, and low CLASS scores were based on the distribution of the dataset values rather than by the categorization of high, medium, and low created by the creators of the CLASS. The numbers and percentages highlighted in bold are those that represent a lack of correspondence between QRIS ratings and CLASS scores (e.g. CLASS score is low but program has a QRIS rating of 3 or higher; CLASS score is high but program has a QRIS rating of a 2).

Table 11b. For programs with each QRIS rating how many had low, medium, and high CLASS scores? (Toddler)

		QRIS Ratings			
CLASS		2(%)	3(%)	4(%)	5(%)
Emotional Support					
	Low	21(47%)	**6(24%)**	**5(42%)**	**8(36%)**
	Medium	15(33%)	14(56%)	5(42%)	14(64%)
	High	**9(20%)**	5(20%)	2(17%)	0(0%)
Instructional Support					
	Low	7(16%)	**0(0%)**	**1(8%)**	**3(14%)**
	Medium	21(47%)	13(52%)	4(33%)	11(50%)
	High	**17(38%)**	12(48%)	7(58%)	8(36%)

2 Toddler Measure does not Include Organized Classrooms. Behavioral Guidance included in Emotional Support

Notes. The cut-off points used to create high, medium, and low CLASS scores were based on the distribution of the dataset values rather than by the categorization of high, medium, and low created by the creators of the CLASS. The numbers and percentages highlighted in bold are those that represent a lack of correspondence between QRIS ratings and CLASS scores (e.g. CLASS score is low but program has a QRIS rating of 3 or higher; CLASS score is high but program has a QRIS rating of a 2).

Table 11c. For programs with each QRIS rating how many had low, medium, and high CLASS scores? (Combined)

		QRIS Ratings			
CLASS		2(%)	3(%)	4(%)	5(%)
Emotional Support (ES)					
	Low	6(32%)	**5(18%)**	**0(0%)**	**3(25%)**
	Medium	8(42%)	17(61%)	2(50%)	5(42%)
	High	**5(26%)**	6(21%)	2(50%)	4(33%)
Organized Classrooms (OC)					
	Low	4(21%)	**5(18%)**	0(0%)	**3(25%)**
	Medium	12(63%)	13(46%)	2(50%)	3(25%)
	High	**3(16%)**	10(36%)	2(50%)	6(50%)
Instructional Support (IS)					
	Low	6(32%)	**4(14%)**	**1(25%)**	**2(17%)**
	Medium	12(63%)	14(50%)	2(50%)	6(50%)
	High	**1(5%)**	10(36%)	1(25%)	4(33%)

Notes. The cut-off points used to create high, medium, and low CLASS scores were based on the distribution of the dataset values rather than by the categorization of high, medium, and low created by the creators of the CLASS. The numbers and percentages highlighted in bold are those that represent a lack of correspondence between QRIS ratings and CLASS scores (e.g. CLASS score is low but program has a QRIS rating of 3 or higher; CLASS score is high but program has a QRIS rating of a 2).

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

We used three complementary analytic approaches to examine how programs' ratings on specific QRIS standards relate to their CLASS scores: 1) cross-tabs of the correspondence between the QRIS ratings on specific standards and CLASS scores 2) Pearson's correlations to examine associations between QRIS ratings on specific standards and CLASS scores, and 3) Analysis of Variance (ANOVA) tests to detect any differences in CLASS scores based on QRIS ratings on specific standards. Correlations assume a linear relationship between variables, such that each increase in a QRIS rating (e.g. from a 2 to a 3 and a 3 to a 4 etc.) is associated with an equal amount of increase in CLASS scores, in a stair-step type fashion. The Analysis of Variance (ANOVA) tests whether there are any differences in CLASS scores across programs with different QRIS ratings. A significant ANOVA test means that there are differences between programs with different ratings, but does not identify which ratings are different from the others (follow up tests have been conducted to examine this where appropriate). Although work that focuses on specific standards is exploratory, by triangulating evidence across these three approaches we gain confidence in the conclusions we draw from the data.

Only programs rated 2 through 5 on the QRIS are included in the analyses because Level 1 programs do not have QRIS ratings. Throughout this section of the report we use Tables to summarize findings from multiple analyses for each QRIS standard, by type of program and CLASS instrument. ***Figures provide illustrative examples; there are too many comparisons to present each one as a figure.***

Learning and Development

Learning and Development 1: Philosophy

The first standard in the Learning and Development (LD1) domain required programs to have a written philosophy that states the programs' values, beliefs, and goals for both children and families. This standard was fairly difficult for programs, compared to other standards, especially for Registered Family providers. Table 12a1 shows that 18% of all programs were rated a 2 on LD1; this was substantially higher (29%) for Registered Family providers. The far right columns in Table 12a also indicate that over 20% of the Registered Family programs that were rated a 2 on LD1, because they failed to meet the star-3 criteria for this standard, had a "high" CLASS score on all three of the CLASS domains, relative to other programs in this study. This indicates somewhat of a mismatch between the LD1 standard and the CLASS, especially for Registered Family providers. A closer look at the data revealed that most of the programs that did not achieve a star-3 rating or higher on LD1 did have a written philosophy for children, but they did not have one for families.

Table 12a2 reveals only one nearly significant correlation between LD1 ratings and CLASS scores, which is for the Organizational domain when considering all types of programs together, but not for any particular type of care on its own. The correlations are actually largest in size for Registered Family providers, although they are not statistically significant, possibly due to a smaller sample size. The results from the ANOVA tests suggest that there are some modest associations between LD1 ratings and CLASS scores for Centers, and also for all programs when considered together as a group. Yet these links between LD1 ratings and CLASS scores are not linear. Table 12a3 shows that generally, programs rated a 2 or a 3 on LD1 score lower on the Instructional and/or Organizational domains of the CLASS than do programs rated 4 or 5 on LD1, but

that there are no differences in CLASS scores between programs rated 2 and 3, or between those rated 4 and 5. Figure 6 provides an illustration of this pattern, using the PreK CLASS for Centers.

In sum, there is some limited evidence for small, significant links between LD1 and CLASS scores for programs overall and for Centers. Evidence is mixed for Registered Family providers.

Table 12a1. Percentage of programs rated 2 on LD1 with high CLASS scores, by program type

					Of the programs rated 2 what % had "high" quality on CLASS?											
	Percent of programs rated 2				All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD1 philosophy	18%	29%	16%	15%	21%	15%	17%	25%	11%	16%	18%	17%	15%	21%	23%	22%

Table 12a2. Correlations among LD1 and CLASS scores by program type

	All Program Types			Certified Centers			Certified Family			Registered Family		
	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD1	.04	.10	**.12**[+]	.05	.13	.10	.03	.03	.15	.12	.22	.34
N	246	246	206	120	120	120	81	81	64	45	45	35

[+] Correlation is nearing significance at the $p < .10$ level (2-tailed)

Table 12a3. Which LD1 ratings significantly differentiate CLASS scores? A summary of results come from various ANOVA tests

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	nd	2-3 vs 4-5	2-3 vs 4-5^{+}	nd	2-3 vs 4-5^{+}	2-3 vs 4-5^{+}	nd	nd	Nd	Nd	nd	nd
PreK	2 vs 4^{+}	2-3 vs 4-5 2 vs 3-5^{+} 2 vs 4^{+} 2 vs 5^{+}	nd	nd	2-3 vs 4-5 3 vs 5	nd	nd	nd	Nd	^	^	^
Toddler	nd	nd	n/a	nd	nd	n/a	^	^	n/a	^	^	n/a
Combined	nd	nd	nd	^	^	^	nd	nd	Nd	^	^	^

Note. All entries (e.g. 2-3 vs 4-5) in this table represent statistically significant differences in CLASS scores between programs with different LD1 ratings (e.g. those rated 2 or 3 vs those rated 4 or 5 for LD1), with the exception that $^{+}$ represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in Appendix E

nd denotes no statistically significant differences.

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Table 12a4. Are there differences in CLASS scores by LD1 rating? Analysis of Variance (ANOVA) results

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	1.06	1.24	1.75	2.03	1.20	2.06	0.48	0.03	0.47	1.96	0.78	1.36
N	246	246	202	120	120	107	81	81	64	45	45	35
PreK	**2.26^{+}**	**4.01****	1.41	1.58	**3.78***	1.94	0.50	0.51	0.61	^	^	^
N	156	156	156	103	103	103	33	33	33			
Toddler	0.57	0.05	n/a	1.33	0.38	n/a	^	^	n/a	^	^	n/a
N	104	104		70	70							
Combined	0.39	0.25	0.18	^	^	^	0.13	0.20	0.02	^	^	^
N	63	63	63				34	34	34			

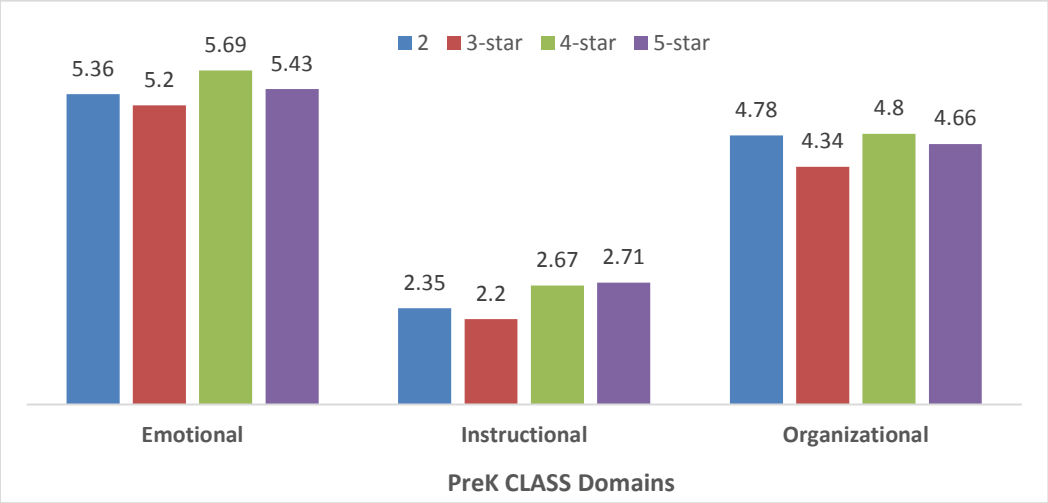
^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

$^{+}$ Correlation is nearing significance at the $p < .10$ level (2-tailed)

* Correlation is significant at the $p < .05$ level (2-tailed)

** Correlation is significant at the $p < .01$ level (2-tailed)

Figure 6. PreK CLASS scores by rating on LD1 (Philosophy) in Centers



Learning and Development 2: Curriculum Use

The Learning and Development 2 (LD2) standard requires programs to use a curriculum to facilitate children's learning and development. There are two ways to meet the standard. Programs can either use a curriculum that is pre-approved, or listed in Oregon's QRIS, or demonstrate how the curriculum the program uses supports children's learning and development in 7 out of 10 specified areas for each age group of children. The areas range from being based on theory and research, balancing adult-directed versus child-directed interactions, includes content areas such as math, science, literacy, and social studies, etc. LD2 is unique in that it only includes criteria for meeting the LD2 standard at the 3-star Level. There are no additional requirements at the 4-star or 5-star levels.

A relatively small percentage (11%) of programs failed to achieve a rating of 3 on LD2, although the percentage was slightly higher at 16% for Registered Family providers (Table 12b1). Among Registered Family providers, over 20% of those rated a 2 on LD had high CLASS scores on Emotional Support and Organizational Support domains. Yet only 8% of these had high CLASS scores on the Instructional Support domain, relative to other programs in this study. This suggests a better alignment between LD2 and the Instructional domain than for the Emotional or Organizational domains of the CLASS for Registered Family providers.

Since all programs are rated either a 2 or a 3 on LD2 we do not report correlations for the links between LD2 and CLASS scores. Results from ANOVA tests examining differences in CLASS scores between programs rated 2 and those rated 3 on LD2 (Table 12b3) reveal a statistically significant difference for the Instructional domain among Certified Family providers (see also Figure 7). The only differences between programs

rated 2 and those rated 3 on LD, even those that are nearing but not quite reaching statistical significance, are seen for the Instructional domain of the CLASS.

Table 12b1. Percentage of programs rated 2 on LD2 with high CLASS scores, by program type

					Of the programs rated 2 what % had "high" quality on CLASS?											
	Percent of programs rated 2				All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD2: curriculum	11%	16%	9%	12%	7%	4%	9%	10%	8%	7%	5%	6%	7%	21%	8%	22%

Table 12b2. Correlations among LD2 (curriculum) and CLASS scores by program type

N/A

Table 12b3.Which LD2 ratings significantly differentiate CLASS scores? A summary of results come from various ANOVA tests

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	nd	2 vs 3-5^{+}	nd	nd	nd	nd	nd	2 vs 3-5	nd	nd	nd	nd
PreK	nd	2 vs 3-5^{+}	nd	nd	2 vs 3-5^{+}	nd	nd	nd	nd	^	^	^
Toddler	nd	nd	n/a	nd	nd	n/a	^	^	n/a	^	^	n/a
Combined	nd	nd	nd	^	^	^	nd	nd	nd	^	^	^

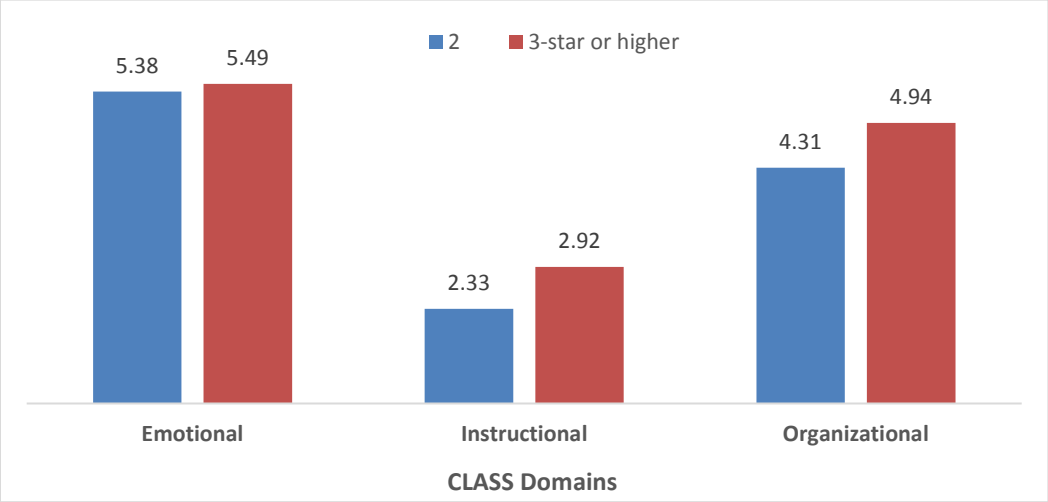
Note. All entries (e.g. 2 vs 3-5) in this table represent statistically significant differences in CLASS scores between programs with different LD2 ratings (e.g. those rated 2 vs those rated 3, 4 or 5 for LD2), with the exception that $^{+}$ represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in the Appendix.

The only comparisons that are meaningful for LD2 are between programs rated 2 vs those rated 3 or higher because LD only has standards for programs at the 3-star level. Nothing additional is required at the 4- or 5-star levels for LD2.

nd denotes no statistically significant differences.

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Figure 7. CLASS Scores by rating on LD2 (Curriculum) in Certified Family programs



Note. CLASS scores are the average of the PreK, Toddler, and Combined instruments
The only comparisons that are meaningful for LD2 are between programs rated 2 vs those rated 3 or higher because LD only has standards for programs at the 3-star Level. Nothing additional is required at the 4- or 5-star levels for LD2.

Supplemental analysis to better understand Links between Curricula Use and observed adult-child interactions (CLASS scores).

The Oregon QRIS Implementation Team requested additional analysis of LD2 to help inform the revision process. They asked whether there was a difference between programs that achieved a rating of 3 on LD2 based on use of a listed or pre-approved curricula versus and those meeting the criteria for 7 out of 10 areas. Analyses revealed only one statistically significant difference. Programs that used a listed curricula had higher Instructional Support scores on the Toddler version of the CLASS than those that met the LD2 standard by passing 7 out of 10 indicators ($F(103) = 4.45$, $p = .03$). There were no differences on CLASS scores based on how many of the 10 indicators programs met. Most of the programs met all 10 indicators.

Conversations with the Oregon QRIS mini review team spurred an interest in exploring the possibility of combining LD2 with another standard, LD7 (focused on programs' use of planned curriculum activities) into one new standard. The Validation Study team used existing data from the information programs submitted for LD2 and LD7 to create a new LD2&7 combined "standard". This is not an actual standard; this was an analytic exercise to explore a possibility. To create the new "standard" the Validation Study team used programs' rating from LD2 as the base; they had to get a rating of 3 showing they used a curriculum to support children's learning and development to receive a score of 3 or higher on the new combined curriculum standard. Then, to create scores of 3 or higher the Validation Study team used the programs' score on LD7 to capture the extent to which they used planned curricula activities, as specified by LD7.

Results revealed that only 11% of programs would have been rated a 2 on the new LD2&7 standard, which is equivalent to LD2, and fewer than the 19% of programs rated a 2 on LD7. More importantly, the new LD2&7 standard had stronger and more consistent links with programs' CLASS scores (Table 12b4) than LD7 alone (correlations were not available for LD2 because programs were rated either 2 or 3; there were no values of 4 or 5 to include in a correlation). Results from the ANOVA tests also point to more consistent links between LD2&7 combined than from LD2 or LD7 alone.

Table 12b4. *Exploring combining LD2 and LD7.* Correlations among a new combined LD2&7 standard and CLASS scores by program type

	All Program Types			Certified Centers			Certified Family			Registered Family		
	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD2&7 combo	.05	**.13***	**.16***	.05	.07	**.12***	.14	**.24***	**.31***	-.08	.12	.04
N	246	246	206	120	120	120	81	81	64	45	45	35

⁺ Correlation is nearing significance at the $p < .10$ level (2-tailed)

* Correlation is significant at the $p < .05$ level (2-tailed)

Table 12b5. *Exploring combining LD2 andLD7*. Which LD2&7combined ratings would significantly differentiate CLASS scores? A summary of results come from various ANOVA tests.

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	nd	2-3 vs 4-5 2 vs 3-5⁺	2-3 vs 4-5	2 vs 4⁺ 3 vs 4⁺	nd	nd	nd	2 vs 3-5	2-3 vs 4-5	nd	nd	nd
PreK	2-3 vs 4-5 2 vs 4 3 vs 4	2 vs 3-5⁺ 2 vs 4⁺ 3 vs 4⁺ 3 vs 5	2-3 vs 4-5	nd	2-3 vs 4-5	nd	2-3 vs 4-5	2-3 vs 4-5⁺	2-3 vs 4-5	^	^	^
Toddler	nd	nd	n/a	nd	nd	n/a	^	^	n/a	^	^	n/a
Combined	nd	nd	nd	^	^	^	nd	nd	nd	^	^	^

Note. All entries (e.g. 2 vs 3-5) in this table represent statistically significant differences in CLASS scores between programs with different LD2&7 combined ratings (e.g. those rated 2 vs those rated 3, 4 or 5 for LD2&7 combined), with the exception that ⁺ represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in the Appendix.

nd denotes no statistically significant differences.

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Learning and Development 3: Appropriate Indoor Environment

Overall, programs were fairly successful in achieving a rating of 3-star or higher on LD3: indoor environments. While both LD3 and LD4 are related to physical environments, which may not be as *directly* related to adult-child interactions as some of the other LD standards may be, we analyzed the relationship of these standards to CLASS scores in order to understand the full set of LD standards. Only 9% of them received a rating of a 2 on this standard. Few of the programs rated a 2 on LD3 had high CLASS scores, relative to other programs in this study (Table 12c1). This indicates good correspondence between the LD3 ratings and the CLASS scores across all three types of programs.

Programs' ratings on LD3 were associated with CLASS scores in the Organization domain for Certified Family providers, and also for all programs overall when considered together as a group (Table 12c2). Programs' LD3 ratings also differentiated CLASS scores in ANOVA tests (Table 12c4). When combining all three types of programs together, LD3 ratings differentiate CLASS scores in each of the three CLASS domains: Emotional, Organizational, and Instructional (see also Figure 8). It is notable that LD3 is linked not only with the overall and PreK CLASS scores, but also with the Toddler CLASS scores on the Emotional and Behavioral Support domain. Few QRIS standards are linked with the Toddler CLASS.

In sum, there is some evidence for small significant associations between LD3 ratings and CLASS scores.

Table 12c1. Percentage of programs rated 2 on LD3 with high CLASS scores, by program type

					Of the programs rated 2 what % had "high" quality on CLASS?											
	Percent of programs rated 2				All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD3: indoor env.	9%	9%	10%	6%	7%	5%	8%	10%	8%	9%	5%	3%	7%	7%	0%	9%

Table 12c2. Correlations among LD3 and CLASS scores by program type

	All Program Types			Certified Centers			Certified Family			Registered Family		
	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD3	.07	**.11**$^{+}$	**.16***	.04	.10	.10	.14	.12	**.27***	.12	.14	.27
N	246	246	206	120	120	120	81	81	64	45	45	35

$^{+}$ Correlation is nearing significance at the $p < .10$ level (2-tailed)

* Correlation is significant at the $p < .05$ level (2-tailed)

Table 12c3. Which LD3 ratings significantly differentiate CLASS scores? A summary of results come from various ANOVA tests

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	2 vs 3-5	2-3 vs 4-5	2-3 vs 4-5	nd	nd	nd	nd	nd	2-3 vs 4-5	2 vs 3-5^{+}	nd	nd
PreK	nd	2-3 vs 4-5 3 vs 4^{+} 3 vs 5^{+}	nd	nd	2-3 vs 4-5 3 vs 5^{+}	nd	nd	2-3 vs 4-5	nd	^	^	^
Toddler	2 vs 3-5	nd	n/a	2 vs 3-5^{+}	nd	n/a	^	^	n/a	^	^	n/a
Combined	nd	nd	nd	^	^	^	nd	nd	nd	^	^	^

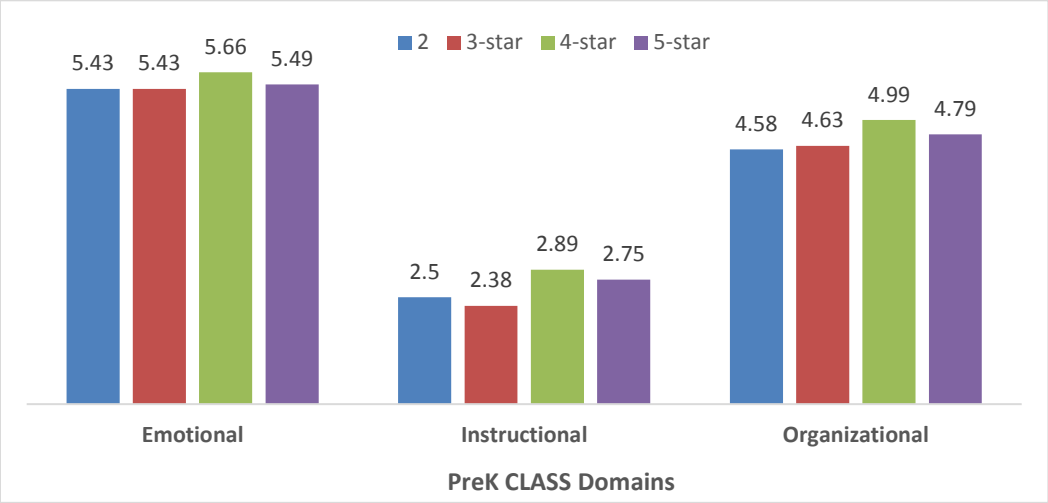
Note. All entries (e.g. 2 vs 3-5) in this table represent statistically significant differences in CLASS scores between programs with different LD3 ratings (e.g. those rated 2 vs those rated 3, 4 or 5 for LD3), with the exception that $^{+}$ represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in the Appendix.
nd denotes no statistically significant differences.
^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Table 12c4. Are there differences in CLASS scores by LD3 rating? Analysis of Variance (ANOVA) results

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	1.89	1.64	**2.76***	0.69	0.51	0.44	0.92	0.54	**2.45+**	1.46	0.71	0.94
N	246	246	202	120	120	107	81	81	64	45	45	35
PreK	0.68	**3.63***	1.23	0.45	**2.36+**	0.16	1.33	1.44	2.01	^	^	^
N	156	156	156	103	103	103	33	33	33			
Toddler	**2.43^{+}**	0.50	n/a	1.74	0.81	n/a	^	^	n/a	^	^	n/a
N	104	104		70	70							
Combined	0.77	0.35	0.81	^	^	^	0.21	0.07	0.16	^	^	^
N	63	63	63				34	34	34			

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).
$^{+}$ Correlation is nearing significance at the $p < .10$ level (2-tailed)
* Correlation is significant at the $p < .05$ level (2-tailed)
** Correlation is significant at the $p < .01$ level (2-tailed)

Figure 8. PreK CLASS scores by rating on LD3 (all program types)



Learning and Development 4: Indoor Furnishings

The Learning and Development 4 standard relates to indoor furnishings; for home-based providers it only pertains to parts of the home made available to children.

A relatively small percentage (11%) of programs failed to achieve a rating of 3 on LD4, although the percentage was slightly higher at 16% for Registered Family providers (Table 12d1). Among all types of programs, fewer than 15% of those rated a 2 on LD4 had high CLASS scores on Emotional Support and Organizational Support domains, relative to other programs in this study. This percentage was lowest overall (8%) for the Instructional Support domain.

The pattern of the correlations presented in Table 12d2, and ANOVAs presented in Table 12d3 suggest that there is a small, significant link between LD4 ratings and CLASS scores. However, this is primarily only the case for Centers. The overall CLASS scores combined across all types of care are linked with CLASS scores in both correlations and ANOVAs, but when examined by program type, only the links for Centers are statistically significant, or even nearing significance.

The links between LD4 ratings and CLASS scores are not linear. Table 12d3 and Figure 9 show that generally, programs rated a 2 or a 3 on LD4 score lower on the Instructional and/or Organizational domains of the CLASS than do programs rated 4 or 5 on LD4, but that there are no differences in CLASS scores between programs rated 2 and 3, or between those rated 4 and 5.

In sum, there is some limited evidence for small, significant links between LD4 and CLASS scores on the Instructional and Organizational domains for programs overall, and for Centers.

Table 12d1. Percentage of programs rated 2 on LD4 with high CLASS scores, by program type

					Of the programs rated 2 what % had "high" quality on CLASS?											
	Percent of programs rated 2				All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD4: Indoor Furnishings	11%	16%	13%	7%	11%	8%	11%	15%	5%	13%	5%	9%	7%	14%	15%	17%

Table 12d2. Correlations among LD4 (indoor furnishings) and CLASS scores by program type

	All Program Types			Certified Centers			Certified Family			Registered Family		
	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD4	.09	**.12**$^{+}$	**.13**$^{+}$	.08	**.18***	**.17**$^{+}$	.14	.08	.14	.10	.04	.08
N	246	246	206	120	120	120	81	81	64	45	45	35

$^{+}$ Correlation is nearing significance at the $p < .10$ level (2-tailed)

* Correlation is significant at the $p < .05$ level (2-tailed)

Table 12d3. Which LD4 ratings significantly differentiate CLASS scores? A summary of results come from various ANOVA tests

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	nd	2-3 vs 4-5^{+}	2-3 vs 4-5^{+}	nd	2 vs 3-5 2-3 vs 4-5^{+}	nd	Nd	nd	nd	nd	nd	nd
PreK	nd	2-3 vs 4-5	nd	nd	nd	nd	Nd	nd	nd	^	^	^
Toddler	nd	nd	n/a	2-3 vs 4-5^{+}	2-3 vs 4-5^{+}	n/a	^	^	n/a	^	^	n/a
Combined	nd	nd	nd	^	^	^	Nd	nd	nd	^	^	^

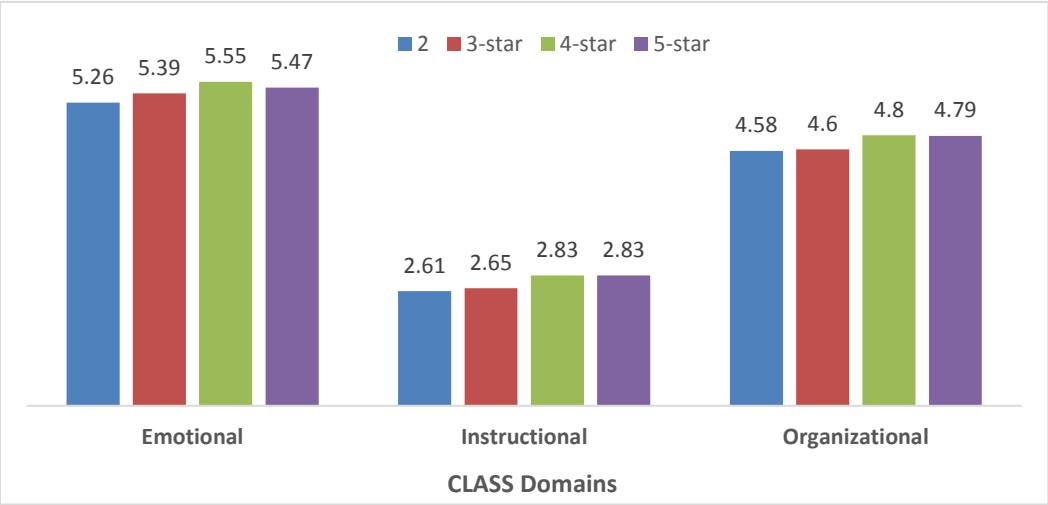
Note. All entries (e.g. 2 vs 3-5) in this table represent statistically significant differences in CLASS scores between programs with different LD4 ratings (e.g. those rated 2 vs those rated 3, 4 or 5 for LD3), with the exception that $^{+}$ represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in the Appendix.
nd denotes no statistically significant differences.
^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Table 12d4. Are there differences in CLASS scores by LD4 rating? Analysis of Variance (ANOVA) results

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	0.98	1.23	1.20	0.67	1.83	1.04	0.70	0.53	0.44	0.95	0.42	0.53
N	246	246	202	120	120	107	81	81	64	45	45	35
PreK	0.97	**2.18+**	0.77	0.57	**2.52+**	0.37	0.83	0.55	0.73	^	^	^
N	156	156	156	103	103	103	33	33	33			
Toddler	0.68	1.28	n/a	1.46	1.49	n/a	^	^	n/a	^	^	n/a
N	104	104		70	70							
Combined	0.89	0.72	0.48	^	^	^	0.07	0.59	0.11	^	^	^
N	63	63	63				34	34	34			

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).
$^{+}$ Correlation is nearing significance at the $p < .10$ level (2-tailed)
* Correlation is significant at the $p < .05$ level (2-tailed)
** Correlation is significant at the $p < .01$ level (2-tailed)

Figure 9. CLASS scores by rating on LD4 (all program types)



Note. CLASS scores are the average of the PreK, Toddler, and Combined instruments

Learning and Development 5: Outdoor Environment

The Learning and Development 5 (LD5) standard includes criteria for gross motor space and materials; again, while these indicators of the quality of physical environments are not necessarily expected to *directly* relate to adult-child interactions, we examined all 12 of the LD standards, due to the conceptual overlap between this domain overall, and the adult-child interactions. A small percentage (7%) of programs fail to meet this standard at the 3-star Level or higher (Table 12e1). Few of these demonstrate high quality on the CLASS, relative to other programs in this study. This is especially the case for Certified Family providers, and for Centers on the Instructional and Organizational domains. Overall, the patterns shown in Table 12e1 indicate that programs with high quality adult-child interactions do not tend to get stuck at a Level 2 on LD5.

LD5 ratings are also linked with CLASS scores for Centers and Certified Family programs, particularly on the Instructional and Organizational domains (Tables 12e2 and 12e3). This is not the case for Registered Family programs. A closer look at data summarized throughout tables 15a-c and Figure 10 reveal that the small number of programs rated 2 on LD5 have substantially lower scores on the Instructional and Organizational domains of the CLASS. This small number of Level 2 programs appears to be primarily responsible for the links between LD5 and CLASS.

In sum, there is some evidence for small, significant links between LD5 and CLASS scores in Centers and Certified Family programs.

Table 12e1. Percentage of programs rated 2 on LD5 with high CLASS scores, by program type

					Of the programs rated 2 what % had "high" quality on CLASS?											
	Percent of programs rated 2				All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD5: Outdoor Environment	7%	11%	6%	5%	11%	5%	8%	20%	5%	2%	0%	0%	9%	14%	15%	17%

Table 12e2. Correlations among LD5 (outdoor environment) and CLASS scores by program type

	All Program Types			Certified Centers			Certified Family			Registered Family		
	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD5	.06	**.14***	**.19****	.03	**.17**$^{+}$**	**.22***	.15	**.23***	**.30***	.02	-.07	-.03
N	246	246	206	120	120	120	81	81	64	45	45	35

Table 12e3. Which LD5 ratings significantly differentiate CLASS scores? A summary of results come from various ANOVA tests

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	**CLASS domains**			**CLASS domains**			**CLASS domains**			**CLASS domains**		
Instrument	**ES**	**IS**	**OC**	**ES**	**IS**	**OC**	**ES**	**IS**	**OC**	**ES**	**IS**	**OC**
Average	nd	2-3 vs 4-5	2-3 vs 4-5	nd	2-3 vs 4-5	2-3 vs 4-5 2 vs 3-5	nd	2 vs 3-5^{+} 2-3 vs 4-5^{+}	2-3 vs 4-5	Nd	nd	nd
PreK	2-3 vs 4-5 3 vs 4^{+}	2 vs 3-5 2-3 vs 4-5 2 vs 4 2 vs 5 3 vs 4 3 vs 5	2 vs 3-5 2-3 vs 4-5	2-3 vs 4-5^{+} 3 vs 4^{+}	2 vs 3-5^{+} 2-3 vs 4-5 3 vs 4^{+} 3 vs 5	2-3 vs 4-5 2 vs 3-5^{+}	2 vs 3-5^{+} 2-3 vs 4-5^{+}	2 vs 3-5^{+} 2-3 vs 4-5^{+}	2 vs 3-5^{+} 2-3 vs 4-5	^	^	^
Toddler	2 vs 3-5$^{+\#}$	nd	n/a	2 vs 3-5$^{\#}$ 2 vs 3$^{\#}$	nd	n/a	^	^	n/a	^	^	n/a
Combined	nd	nd	nd	^	^	^	nd	nd	Nd	^	^	^

Note. All entries (e.g. 2 vs 3-5) in this table represent statistically significant differences in CLASS scores between programs with different LD5 ratings (e.g. those rated 2 vs those rated 3, 4 or 5 for LD5), with the exception that $^{+}$ represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in the Appendix.

significant differences is in the opposite direction than expected; programs with ratings of 2 on LD5 have *higher* CLASS scores on the Emotional and Behavioral Support domain of the Toddler CLASS than those with LD5 ratings of 3.

nd denotes no statistically significant differences.

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Table 12e4. Are there differences in CLASS scores by LD5 rating? Analysis of Variance (ANOVA) results

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS Instrument	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	1.13	1.74	**2.60+**	1.99	1.26	**2.47+**	0.62	1.86	2.06	0.20	0.40	0.52
N	246	246	202	120	120	107	81	81	64	45	45	35
PreK	**3.05***	**6.89****	**2.55+**	**2.48+**	**6.09****	**2.43+**	1.84	1.86	2.10	^	^	^
N	156	156	156	103	103	103	33	33	33			
Toddler	1.87	0.32	n/a	**3.43*#**	0.60	n/a	^	^	n/a	^	^	n/a
N	104	104		70	70							
Combined	0.14	0.18	0.46	^	^	^	0.12	0.10	0.12	^	^	^
N	63	63	63				34	34	34			

significant differences is in the opposite direction than expected; programs with ratings of 2 on LD5 have *higher* CLASS scores on the Emotional and Behavioral Support domain of the Toddler CLASS than those with LD5 ratings of 3.

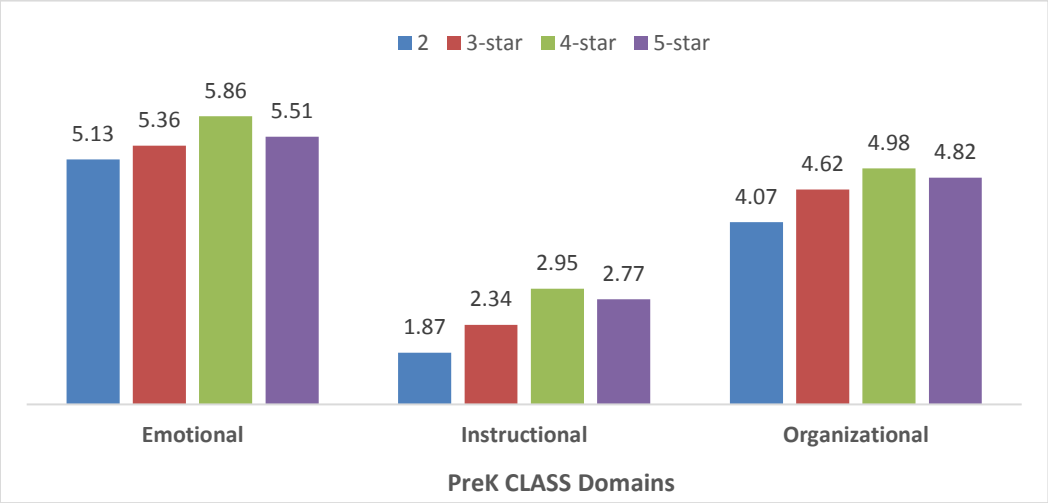
^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

+ Correlation is nearing significance at the $p < .10$ level (2-tailed)

* Correlation is significant at the $p < .05$ level (2-tailed)

** Correlation is significant at the $p < .01$ level (2-tailed)

Figure 10. PreK CLASS scores by rating on LD5 (all program types)



Learning and Development 6: Materials Support Development

The Learning and Development 6 (LD6) standard outlines how programs are expected to use materials to support children's learning and development. Overall, 11% of programs fail to meet this standard at the 3-star Level or higher (Table 12f1). Few of these programs rated a 2 on LD6 demonstrate high quality on the CLASS, relative to other programs in this study, with the exception of the Emotional Support domain for Registered Family providers (21%).

The pattern of the correlations presented in Table 12f2, and ANOVAs presented in Table 12f3 suggest that there is a small, significant link between LD6 ratings and CLASS scores. The CLASS domain that is most consistently linked with LD6 ratings across the three program types is Organized Classrooms, which makes sense because this standard is focused on using materials to support children's learning. Although the correlations in Table 12f2 are only statistically significant for Certified Family programs, the size of the correlation for Registered Family providers (.19) is indicative of a possible link that could have been limited in significance due to fewer Registered Family providers participating in the study.

Table 12f3 also indicates a non-linear link between LD6 ratings and CLASS scores for both Centers and Certified Family providers. Programs rated a 2 or a 3 on LD6 score lower on the Instructional, Organizational, and/or Emotional (Certified Family only) domains of the CLASS than do programs rated 4 or 5 on LD6.

In sum, there is some evidence for small, significant links between LD6 and CLASS scores in Centers and Certified Family programs.

Table 12f1. Percentage of programs rated 2 on LD6 with high CLASS scores, by program type

					Of the programs rated 2 what % had "high" quality on CLASS?											
	Percent of programs rated 2				All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD6: Materials Support Dev.	11%	13%	13%	6%	13%	6%	11%	15%	8%	16%	5%	3%	4%	21%	8%	17%

Table 12f2. Correlations among LD6 (materials support development) and CLASS scores by program type

	All Program Types			Certified Centers			Certified Family			Registered Family		
	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD6	.05	.10	**.15***	.04	.10	.13	.11	.11	**.23**$^{+}$	.07	.07	.19
N	246	246	206	120	120	120	81	81	64	45	45	35

$^{+}$ Correlation is nearing significance at the $p < .10$ level (2-tailed)

* Correlation is significant at the $p < .05$ level (2-tailed)

Table 12f3. Which LD6 ratings significantly differentiate CLASS scores? A summary of results come from various ANOVA tests

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	nd	2-3 vs 4-5^{+}	2-3 vs 4-5^{+}	nd	nd	2-3 vs 4-5^{+}	nd	nd	2-3 vs 4-5^{+}		nd	nd
PreK	nd	2-3 vs 4-5 3 vs 4 3 vs 5^{+}	2-3 vs 4-5^{+}	nd	2-3 vs 4-5 3 vs 5	nd	2 vs 3-5	nd	nd	^	^	^
Toddler	nd	nd	n/a	nd	nd	n/a	^	^	n/a	^	^	n/a
Combined	nd	nd	nd	^	^	^	nd	nd	nd	^	^	^

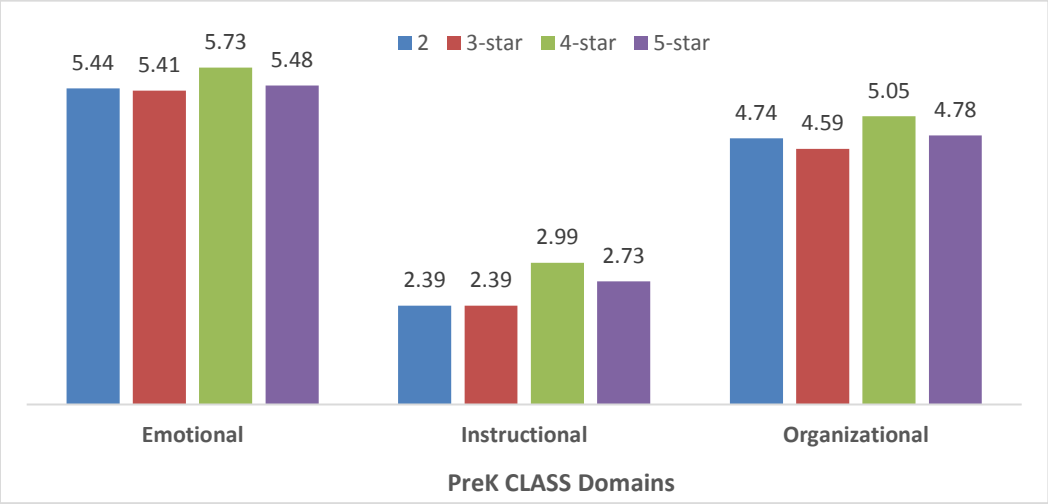
Note. All entries (e.g. 2 vs 3-5) in this table represent statistically significant differences in CLASS scores between programs with different LD6 ratings (e.g. those rated 2 vs those rated 3, 4 or 5 for LD6), with the exception that $^{+}$ represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in the Appendix.
nd denotes no statistically significant differences.
^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Table 12f4. Are there differences in CLASS scores by LD6 rating? Analysis of Variance (ANOVA) results

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	0.58	1.32	**2.64+**	0.61	0.79	1.16	1.31	0.33	1.36	0.86	0.83	1.08
N	246	246	202	120	120	107	81	81	64	45	45	35
PreK	1.11	**4.21****	1.63	0.98	**3.48***	2.00	**3.14***	0.67	1.27	^	^	^
N	156	156	156	103	103	103	33	33	33			
Toddler	0.48	0.53	n/a	1.41	0.85	n/a	^	^	n/a	^	^	n/a
N	104	104		70	70							
Combined	0.26	0.39	0.68	^	^	^	0.60	0.15	0.03	^	^	^
N	63	63	63				34	34	34			

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).
$^{+}$ Correlation is nearing significance at the $p < .10$ level (2-tailed)
* Correlation is significant at the $p < .05$ level (2-tailed)
** Correlation is significant at the $p < .01$ level (2-tailed)

Figure 11. PreK CLASS scores by rating on LD6 (all program types)



Learning and Development 7: Planned Curriculum Activities

The Learning and Development 7 (LD7) standard outlines how programs are expected to use planned curriculum activities to support children's learning and development. This was a fairly difficult standard for programs to meet, evidenced by 19% of programs failing to meet this standard at the 3-star Level or higher (Table 12g1). Within both Centers and Registered Family programs approximately 20% or more of these programs that were rated 2 on LD7 showed high quality on the CLASS, relative to other programs in this study (Table 12g1). This indicates a somewhat of a mismatch between LD7 and CLASS scores.

Tables 12g2 and 12g3, and Figure 12, show some small significant links between LD7 ratings and CLASS scores, but these links are not as strong or as consistent as those presented earlier for the "new" combined LD2&7 standard (see the LD2 section, above).

In sum, there is some evidence for small, significant links between LD7 and CLASS scores but they are not as strong or consistent as those for what could be a "new" standard combining LD2&7, which both address curricula.

Table 12g1. Percentage of programs rated 2 on LD7 with high CLASS scores by program type

					Of the programs rated 2 what % had "high" quality on CLASS?											
	Percent of programs rated 2				All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD7: Planned Curriculum	19%	31%	19%	11%	18%	15%	17%	25%	19%	18%	5%	9%	11%	29%	235	26%

Table 12g2. Correlations among LD7 (planned curriculum activities) and CLASS scores by program type

	All Program Types			Certified Centers			Certified Family			Registered Family		
	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD7	.04	**.12⁺**	**.15***	-.00	.07	.07	.12	.18	**.27***	.06	.14	.20
N	246	246	206	120	120	120	81	81	64	45	45	35

⁺ Correlation is nearing significance at the $p < .10$ level (2-tailed)

* Correlation is significant at the $p < .05$ level (2-tailed)

Table 12g3. Which LD7 ratings significantly differentiate CLASS scores? A summary of results come from various ANOVA tests

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	nd	nd	3 vs 5+	nd	nd	nd	nd	nd	2-3 vs 4-5	nd	nd	nd
PreK	nd	2-3 vs 4-5+ 2 vs 5+ 3 vs 4+ 3 vs 5	nd	nd	2-3 vs 4-5	nd	2 vs 3-5+ 2-3 vs 4-5+	2-3 vs 4-5+	2-3 vs 4-5	^	^	^
Toddler	nd	nd	n/a	3 vs 4	nd	n/a	^	^	n/a	^	^	n/a
Combined	nd	nd	nd	^	^	^	Nd	nd	nd	^	^	^

Note. All entries (e.g. 2 vs 3-5) in this table represent statistically significant differences in CLASS scores between programs with different LD7 ratings (e.g. those rated 2 vs those rated 3, 4 or 5 for LD7), with the exception that ⁺ represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in the Appendix.

nd denotes no statistically significant differences.

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Table 12g4. Are there differences in CLASS scores by LD7 rating? Analysis of Variance (ANOVA) results

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	**CLASS domains**			**CLASS domains**			**CLASS domains**			**CLASS domains**		
Instrument	**ES**	**IS**	**OC**	**ES**	**IS**	**OC**	**ES**	**IS**	**OC**	**ES**	**IS**	**OC**
Average	1.47	1.50	**2.93***	1.99	0.52	0.93	0.42	0.84	2.06	0.03	1.56	2.77
N	246	246	202	120	120	107	81	81	64	45	45	35
PreK	1.98	**4.39****	**2.41+**	1.96	**2.03+**	1.19	1.90	1.13	**2.61+**	^	^	^
N	156	156	156	103	103	103	33	33	33			
Toddler	1.17	0.61	n/a	**3.27***	0.40	n/a	^	^	n/a	^	^	n/a
N	104	104		70	70							
Combined	0.32	0.30	0.26	^	^	^	0.44	0.33	0.02	^	^	^
N	63	63	63				34	34	34			

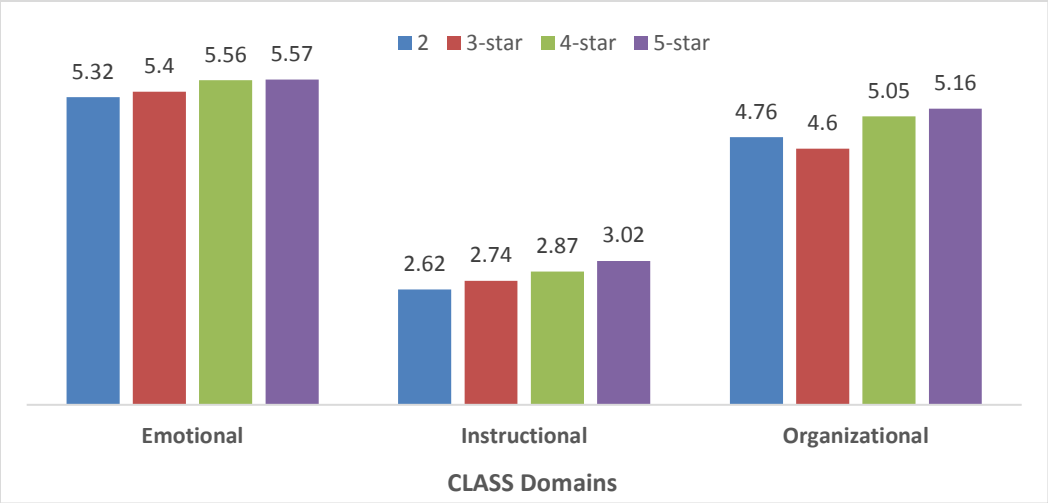
^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

+ Correlation is nearing significance at the $p < .10$ level (2-tailed)

* Correlation is significant at the $p < .05$ level (2-tailed)

** Correlation is significant at the $p < .01$ level (2-tailed)

Figure 12. CLASS scores by rating on LD7 in Certified Family programs



Note. CLASS scores are the average of the PreK, Toddler, and Combined instruments

Learning and Development 8: Routines

The Learning and Development 8 (LD8) standard is based on programs' use of routines to support children's learning and development. Several of the indicators within this LD8 standard specify practices related to child–directed activities, and adult-child interactions. A small percentage (9%) of programs fail to meet this standard at the 3-star Level or higher (Table 12h1). Few of these demonstrate high quality on the CLASS, relative to other programs in this study. This is especially the case for Centers and Certified Family providers. Overall, the patterns shown in Table 12h1 indicate that programs with high quality adult-child interactions, relative to other programs in this study, do not tend to get stuck at a Level 2 on LD8.

LD8 ratings are also linked with CLASS scores on the Instructional and Organizational domains for programs overall (Table 12h2). Although correlations are less consistently significant statistically for each type of program considered separately (and not even *nearing* statistical significance for Registered Family), the pattern of findings (positive correlations ranging from $r = .16$ to $r = .26$) is consistent across type of care. Results from ANOVA tests (Table 12h3) are similar, but also point to differences in CLASS scores by LD8 rating on the Emotional Support domain of the CLASS. Table 12h3 also points to differences in the Toddler CLASS instrument based on LD8 ratings (see also Figure 13). This is particularly noteworthy given that few standards appear to differentiate Toddler CLASS scores.

In sum, there is consistent evidence for small, significant links between LD8 and CLASS scores, especially for Centers and Certified Family programs. It is noteworthy that LD8 ratings are linked with all three CLASS domains, and with both the Toddler and PreK CLASS instruments. The pattern of associations among LD8 and CLASS scores are also relatively consistent across all three types of care, although they do not reach statistical significance for Registered Family providers.

Table 12h1. Percentage of programs rated 2 on LD8 with high CLASS scores by program type

					Of the programs rated 2 what % had "high" quality on CLASS?											
	Percent of programs rated 2				All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD8: Routines Support I&D	9	13%	9%	6%	4%	2%	7%	0	5%	7%	0	0	4%	14%	0	13%

Table 12h2. Correlations among LD8 (routines) and CLASS scores by program type

	All Program Types			Certified Centers			Certified Family			Registered Family		
	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	O
LD8	.10	**.17****	**.20****	.13	**.16⁺**	**.19⁺**	.14	**.19⁺**	**.27***	.08	.23	.26
N	246	246	206	120	120	120	81	81	64	45	45	35

⁺ Correlation is nearing significance at the $p < .10$ level (2-tailed)

* Correlation is significant at the $p < .05$ level (2-tailed)

Table 12h3. Which LD8 ratings significantly differentiate CLASS scores? A summary of results come from various ANOVA tests

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	2 vs 3-5 2 vs 4	2 vs 3-5+ 2-3 vs 4-5 2 vs 5	2 vs 3-5 2-3 vs 4-5	nd	nd	2-3 vs 4-5+	nd	nd	2-3 vs 4-5	nd	nd	nd
PreK	2-3 vs 4-5	2-3 vs 4-5 3 vs 5	2-3 vs 4-5	2-3 vs 4-5+	2-3 vs 4-5 3 vs 5	nd	2 vs 3-5 2 vs 4 2 vs 5	nd	2-3 vs 4-5	^	^	^
Toddler	nd	2 vs 3-5	n/a	2 vs 3-5 2 vs 4	nd	n/a	^	^	n/a	^	^	n/a
Combined	nd	nd	nd	^	^	^	nd	nd	nd	^	^	^

Note. All entries (e.g. 2 vs 3-5) in this table represent statistically significant differences in CLASS scores between programs with different LD8 ratings (e.g. those rated 2 vs those rated 3, 4 or 5 for LD8), with the exception that ⁺ represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in the Appendix.

nd denotes no statistically significant differences.

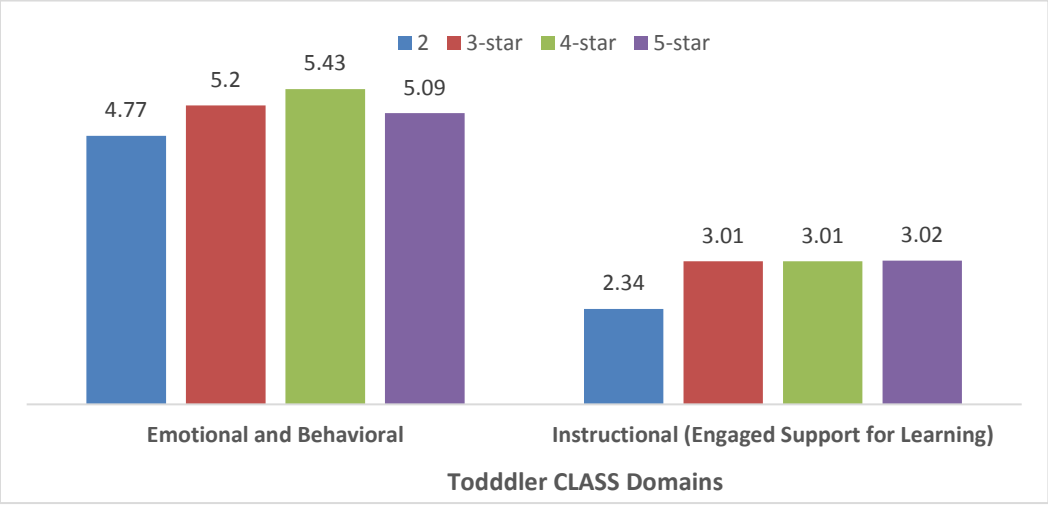
^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Table 12h4. Are there differences in CLASS scores by LD8 rating? Analysis of Variance (ANOVA) results

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	**CLASS domains**			**CLASS domains**			**CLASS domains**			**CLASS domains**		
Instrument	**ES**	**IS**	**OC**	**ES**	**IS**	**OC**	**ES**	**IS**	**OC**	**ES**	**IS**	**OC**
Average	**2.78***	**3.01***	**3.23***	1.64	1.22	1.40	0.67	1.37	1.93	1.38	1.03	0.91
N	246	246	202	120	120	107	81	81	64	45	45	35
PreK	**2.32+**	**4.06***	1.46	1.15	**3.39***	0.79	**4.45***	1.37	1.77	^	^	^
N	156	156	156	103	103	103	33	33	33			
Toddler	1.32	**2.18+**	n/a	**3.92***	0.96	n/a	^	^	n/a	^	^	n/a
N	104	104		70	70							
Combined	1.10	1.02	1.20	^	^	^	0.06	0.11	0.11	^	^	^
N	63	63	63				34	34	34			

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).
+ Correlation is nearing significance at the $p < .10$ level (2-tailed)
* Correlation is significant at the $p < .05$ level (2-tailed)
** Correlation is significant at the $p < .01$ level (2-tailed)

Figure 13. Toddler CLASS scores by rating on LD8 (all program types)



Learning and Development 9: Screening and Assessment

The Learning and Development 9 (LD9) standard includes indicators related to routine screenings of all children at the 3-star Level, and using assessments to plan learning activities for children at the 4- and 5-star levels. LD9 was an extremely difficult standard for programs to meet, evidenced by 42% of programs failing to meet this standard at the 3-star Level or higher (Table 12i1). Across all three program types, high percentages of programs that were rated a 2 on LD9 showed high quality CLASS scores relative to other programs in this study (Table 12i1). This indicates a mismatch between LD9 and CLASS scores. Tables 12i2 and 12i3 support this idea, showing no links between LD9 ratings and CLASS scores for Centers and Certified Family programs (see also Figure 14).

Yet, for Registered Family providers there does appear to be a link between LD9 ratings and CLASS scores (Tables 12i2 and 12i3, and Figure 15). This is important given that few standards are linked statistically with CLASS scores for Registered Family providers.

In sum, there are substantial concerns about LD9 across all three types of programs. Yet, it is also noteworthy that there is some evidence for significant links between LD9 ratings and CLASS scores for Registered Family providers. Further analysis, and/or discussion with the QRIS process evaluation team about LD9 for Registered Family providers may help to illuminate factors that could be relevant for potential revision of Oregon's QRIS.

Table 12i1. Percentage of programs rated 2 on LD9 with high CLASS scores, by program type

					Of the programs rated 2 what % had "high" quality on CLASS?											
	Percent of programs rated 2				All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD9: Screening	42%	51%	42%	38%	39%	29%	37%	40%	27%	38%	36%	34%	37%	43%	23%	35%

Table 12i2. Correlations among LD9 (screening and assessment) and CLASS scores by program type

	All Program Types			Certified Centers			Certified Family			Registered Family		
	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD9	.01	.05	.04	.01	.03	-.00	-.01	-.03	.07	.12	**.36***	**.30⁺**
N	246	246	206	120	120	120	81	81	64	45	45	35

⁺ Correlation is nearing significance at the $p < .10$ level (2-tailed)

* Correlation is significant at the $p < .05$ level (2-tailed)

Table 12i3. Which LD9 ratings significantly differentiate CLASS scores? A summary of results come from various ANOVA tests

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	2 vs 3-5+ 2 vs 3	2 vs 3-5 2 vs 3	nd	nd	nd	nd	nd	nd	nd	2 vs 3-5	2 vs 3-5	2 vs 3-5
PreK	nd	nd	nd	nd	nd	nd	nd		nd	^	^	^
Toddler	nd	nd	n/a	nd	nd	n/a	^	^	n/a	^	^	n/a
Combined	2 vs 3-5 2 vs 3	nd	2 vs 3-5+	^	^	^	nd	nd	nd	^	^	^

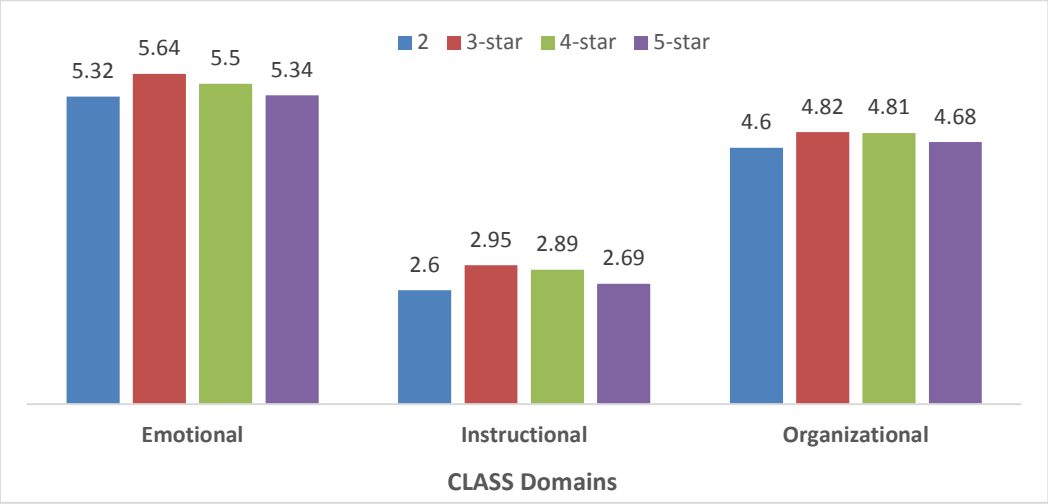
Note. All entries (e.g. 2 vs 3-5) in this table represent statistically significant differences in CLASS scores between programs with different LD9 ratings (e.g. those rated 2 vs those rated 3, 4 or 5 for LD9), with the exception that $^{+}$ represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in the Appendix.
nd denotes no statistically significant differences.
^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Table 12i4. Are there differences in CLASS scores by LD9 rating? Analysis of Variance (ANOVA) results

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	**3.22***	**3.06***	0.98	1.84	1.20	0.11	0.15	0.10	0.17	**5.09***	**3.60***	1.42
N	246	246	202	120	120	107	81	81	64	45	45	35
PreK	1.80	0.56	0.48	1.14	1.09	0.36	0.76	0.40	0.12	^	^	^
N	156	156	156	103	103	103	33	33	33			
Toddler	1.23	1.77	n/a	1.75	0.83	n/a	^	^	n/a	^	^	n/a
N	104	104		70	70							
Combined	**4.36****	1.06	1.75	^	^	^	**2.71+**	0.65	0.92	^	^	^
N	63	63	63				34	34	34			

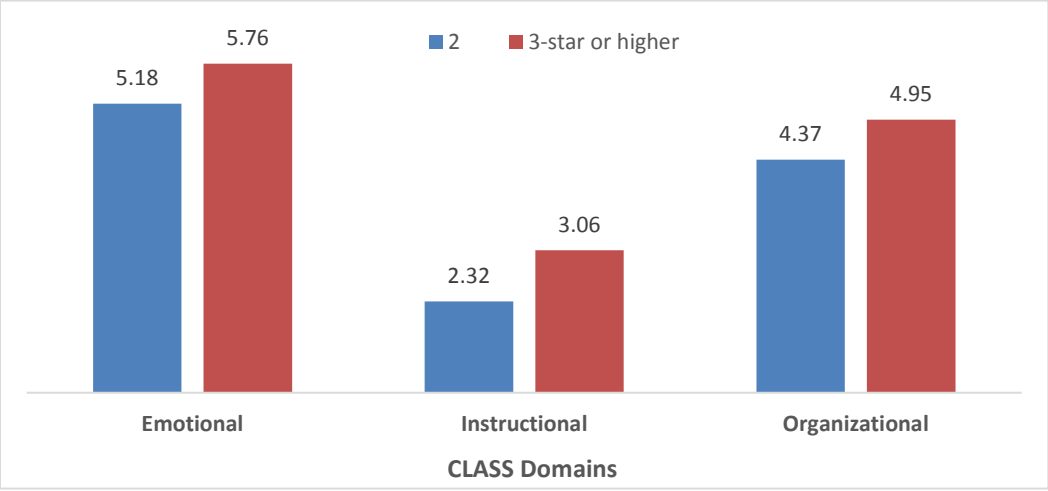
^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).
$^{+}$ Correlation is nearing significance at the $p < .10$ level (2-tailed)
* Correlation is significant at the $p < .05$ level (2-tailed)
** Correlation is significant at the $p < .01$ level (2-tailed)

Figure 14. CLASS scores by rating on LD9 (all program types)



Note. CLASS scores are the average of the PreK, Toddler, and Combined instruments

Figure 15. CLASS scores by rating on LD9 for Registered Family programs



Note. CLASS scores are the average of the PreK, Toddler, and Combined instruments

Learning and Development 10: Group Size, Ratio, and Staffing Patterns

The Learning and Development 10 (LD10) standard is based on staffing patterns (e.g. children have a consistent caregiver throughout the day), group sizes and ratios of children to adults. For Certified and Registered Family programs there is no 5-star indicator for LD10; if they are submitting an overall QRIS portfolio for a 5-star they automatically receive a 5-star rating on LD10 if they pass the 3- and 4-star indicators for LD10. A small percentage (6%) of programs fail to meet this standard at the 3-star Level or higher (Table 12j1). Few of these demonstrate high quality on the CLASS, relative to other programs in this study. Overall, the patterns shown in Table 20a indicate that programs with high quality adult-child interactions, relative to other programs in this study, do not tend to get stuck at a Level 2 on LD10.

LD10 ratings are also linked with CLASS scores on the Instructional and Organizational domains for programs overall (Table 12j2). The only statistically significant correlation for any one type of program considered separately is for the Organizational domain within Certified Family programs. Yet, Registered Family providers also show a correlation among LD10 ratings and CLASS scores for the Instructional and Organizational domains but they are not statistically significant (Table 12j2), likely due to the smaller sample size of Registered Family providers. For Centers, the links between LD10 ratings and CLASS scores appear to be non-linear, with CLASS scores for programs rated 3-star being either equivalent to, or even lower than, CLASS scores for Level 2 programs (Table 12j3). This pattern is shown in Figure 16, for all types of programs overall. This pattern suggests that the 3-star indicator (children have a consistent caregiver throughout the day) may not align well with observed adult-child interactions (CLASS scores) for programs overall. Another complicating factor is that there is no 5-star indicator of the LD10 standard for Certified and Registered Family programs to meet. This also likely leads the overall mixed findings for links between LD10 and CLASS scores.

In sum, there is mixed evidence regarding links between LD10 and CLASS scores.

Table 12j1. Percentage of programs rated 2 on LD10 with high CLASS scores by program type

	Percent of programs rated 2				Of the programs rated 2 what % had "high" quality on CLASS?											
					All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD10: Group Size/ Ratio	6%	4%	8%	4%	4%	2%	4%	10%	5%	4%	0%	0%	2%	0%	0%	4%

Table 12j2. Correlations among LD10 (group size/ratio/staffing) and CLASS scores by program type

	All Program Types			Certified Centers			Certified Family			Registered Family		
	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD10	.07	**.14***	**.16***	.03	.10	.11	**.21⁺**	**.21⁺**	**.27***	.06	.19	.27
N	246	246	206	120	120	120	81	81	64	45	45	35

⁺ Correlation is nearing significance at the $p < .10$ level (2-tailed)

* Correlation is significant at the $p < .05$ level (2-tailed)

Table 12j3. Which LD10 ratings significantly differentiate CLASS scores? A summary of results come from various ANOVA tests

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	nd	2-3 vs 4-5⁺	2-3 vs 4-5	nd	nd	2-3 vs 4-5+	nd	nd	2-3 vs 4-5	nd	nd	nd
PreK	2-3 vs 4-5 3 vs 4	2-3 vs 4-5 3 vs 5	2-3 vs 4-5	2-3 vs 4-5 2 vs 3# 3 vs 4 3 vs 5	2-3 vs 4-5 3 vs 5	nd	nd	nd	nd	^	^	^
Toddler	nd	nd	n/a	nd	nd	n/a	^	^	n/a	^	^	n/a
Combined	nd	nd	nd	^	^	^	Nd	nd	nd	^	^	^

Note. All entries (e.g. 2 vs 3-5) in this table represent statistically significant differences in CLASS scores between programs with different LD10 ratings (e.g. those rated 2 vs those rated 3, 4 or 5 for LD10), with the exception that ⁺ represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in the Appendix.

significant differences is in the opposite direction than expected; programs with ratings of 2 on LD10 have *higher* CLASS scores on the Emotional Support domain of the PreK CLASS than those with LD10 ratings of 3.

nd denotes no statistically significant differences.

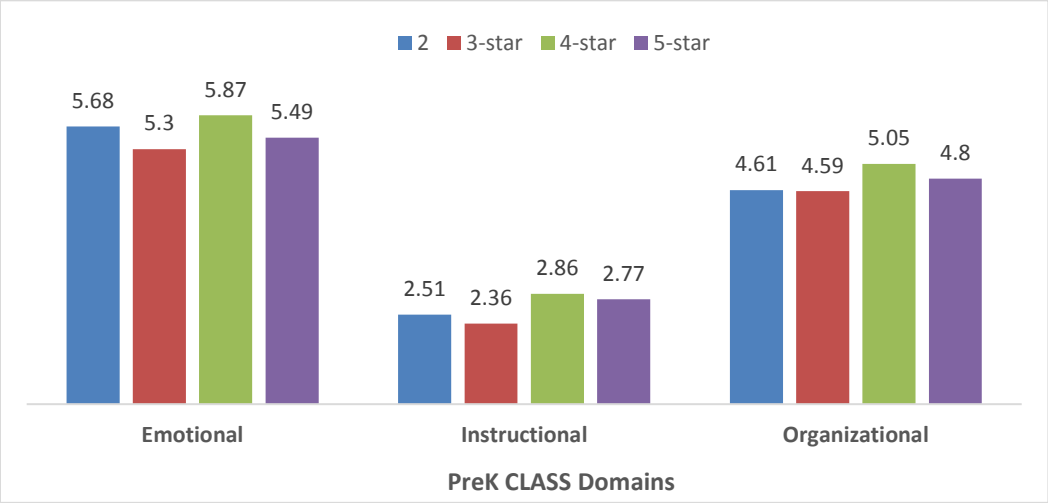
^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Table 12j4. Are there differences in CLASS scores by LD10 rating? Analysis of Variance (ANOVA) results

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	0.85	1.68	**2.69***	1.24	0.78	1.51	1.46	1.59	1.82	1.58	0.62	1.03
N	246	246	202	120	120	107	81	81	64	45	45	35
PreK	**4.22***	3.79	**1.92***	**5.59****	**4.86****	1.58	0.71	0.19	1.30	^	^	^
N	156	156	156	103	103	103	33	33	33			
Toddler	0.98	0.35	n/a	0.54	0.33	n/a	^	^	n/a	^	^	n/a
N	104	104		70	70							
Combined	0.58	0.58	0.30	^	^	^	0.06	0.69	0.24	^	^	^
N	63	63	63				34	34	34			

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).
$^{+}$ Correlation is nearing significance at the $p < .10$ level (2-tailed)
* Correlation is significant at the $p < .05$ level (2-tailed)
** Correlation is significant at the $p < .01$ level (2-tailed)

Figure 16. PreK CLASS scores by rating on LD10 (all program types)



Learning and Development 11: Adult Child Interactions

The Learning and Development 11 (LD11) standard addresses the topic of adult-child interactions. The indicators at the 3- and 4-star levels of the LD11 standard specify that programs must have written guidelines for adult-child interactions that support children's social and emotional needs (3-star Level) and learning, language, and concept development (4-star Level). The 5-star Level involves on onsite observation of the adult-child interaction using the Classroom Assessment Scoring System (CLASS), conducted by the QRIS team (not the Validation Study team).

Overall, 28% of programs fail to meet the 3-star Level or higher on D11 (Table 12k1). The percent is highest in Registered Family programs (42%), followed by Certified Family (30%) and Certified Centers (22%). Substantial proportions of these programs that are rated a 2 on LD11 have CLASS scores that are high, relative to other programs in this study (Table 12k1).

Analysis of linear correlations between LD11 ratings and CLASS scores did not reveal any statistically significant associations (Table 12k2). Yet, results from ANOVA tests did reveal a few significant differences on the CLASS by LD11 rating (Table 12k3). These differences varied by program type. Among Centers and Certified Family programs, those programs who achieved a 5-star rating on LD11 (by scoring a 5 or higher on the CLASS) also showed higher CLASS scores in some domains and/or instruments in the Validation Study than did programs rated 2, 3, or 4 on LD11 (Table 12k3). This was most consistent for the Organizational domain when looking across types of programs overall, although there was also a significant difference in the PreK Instructional domain in Centers and the Emotional domain for Certified Family programs. Given that both the Validation Study and LD11 used the CLASS measure, it is surprising that the 5-star LD11 indicator and CLASS scores from the Validation Study were not even better aligned. One would expect most of the "nd" entries in Table 12k3 (signifying no difference in CLASS scores by LD11 rating) to instead depict differences between programs rated 5 and those rated 2, 3, and/or 4.

This lack of consistent differences may result from the QRIS rating structure and process in which only the programs that submitted for a 5-star QRIS rating overall, and were found to meet the 5-star standards throughout the portfolio, were observed with the CLASS by the QRIS for a chance to achieve a 5-star rating on LD11 and the overall QRIS. Table 12k1 illustrates the challenges associated with this approach. Overall, 26-36 percent of programs that were rated a 2 on LD11, because they failed to provide sufficient written guidelines about adult-child interactions, also had relatively high CLASS scores in the Validation Study. These findings suggest that written guidelines about adult-child interactions do not represent the quality of the interactions that actually take place between adults and children, measured by the CLASS, at least not within Centers and Certified Family programs. For illustrations of these patterns see Figures 17-19.

Additionally, the specific requirements on the QRIS CLASS observation to receive an LD rating of a 5 may be partly responsible for the lack of congruence between LD11 and Validation Study CLASS scores. For example, to receive an LD11 rating of a 5, programs must have an **overall** score of a 5 on the CLASS, averaged across the CLASS domains. They can reach this average with any combination of scores on the 2 (Toddler CLASS) or 3 (PreK CLASS) domains. Given that the average score on the Emotional domain in the Validation Study was higher than a 5 for even the Level 1 programs, it is not surprising that we did not detect associations between LD11 ratings of a 5 and CLASS scores on the Emotional domain. On the other end of the spectrum, the Instructional scores on the CLASS were so low in this Validation study (which is consistent with

other studies in the field), that nearly all of the programs had to leverage higher Emotional and/or Organizational scores in order to bring their total average score up to a 5. The results in Tables 12k2 and 12k3 (overall program rows) indicate that ***this LD11 requirement for an average score of a 5 most commonly resulted in higher Organizational scores***.

For Registered Family providers, there is some indication that LD11 ratings at the 3- and 4- star levels might be at least somewhat better aligned with observed adult-child interactions in the Instructional domain than they are in Centers or Certified Family programs. Registered Family programs rated 3-5 on LD11 had significantly higher Instructional Support scores on the CLASS than did Registered Family programs rated a 2 on LD11 (Table 12k3 and Figure 20). There were also noticeable differences on the Emotional and Organizational domains for Registered Family programs that were not statistically significant, perhaps due in part to the small sample of Registered Family programs (Figure 20). Most (79%) of these Registered Family programs rated 3-5 were rated a 3. This suggests that when Registered Family providers have written guidelines about adult-child interactions they may actually interact with children in ways that reflect higher quality, as defined by the CLASS. In Centers and Certified Family programs, in which the person writing guidelines is often someone other than the one(s) interacting with the children during the CLASS observation, there is no link between written guidelines and observed adult-child interactions.

In Centers and Certified Family programs, in which the person writing guidelines is often someone other than the one(s) interacting with the children during the CLASS observation, there is no link between written guidelines and observed adult-child interactions.

In sum, for Centers and Certified Family programs, only the 5-star indicator (observations) of LD11 ratings are significantly linked with CLASS scores. This is concerning given that many programs get stuck at the 2-star Level on LD11, even many high quality programs. For Registered Family programs there appears to be a link between the various LD11 ratings (3-star and higher) and CLASS scores, especially for Instructional Support. This is particularly notable in light of the fact that with a small sample size for Registered Family programs, very few QRIS standards are significantly associated with CLASS scores.

Table 12k1. Percentage of programs rated 2 on LD11 with high CLASS scores by program type

	Percent of programs rated 2				Of the programs rated 2 what % had "high" quality on CLASS?											
					All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD11: Adult Child Interactions	28%	42%	22%	30%	36%	26%	26%	36%	16%	30%	35%	37%	22%	36%	23%	26%

Table 12k2. Correlations among LD11 and CLASS scores by program type

	All Program Types			Certified Centers			Certified Family			Registered Family		
	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD11	-.01	.07	**.12**$^{+}$	.00	.08	.11	-.04	-.04	.14	.07	**.29**$^{+}$	.17
N	246	246	206	120	120	120	81	81	64	45	45	35

$^{+}$ Correlation is nearing significance at the $p < .10$ level (2-tailed)

* Correlation is significant at the $p < .05$ level (2-tailed)

Table 12k3. Which LD11 ratings significantly differentiate CLASS scores? A summary of results come from various ANOVA tests

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	nd	nd	3 vs 5 4 vs 5	nd	nd	3 vs 5	2 vs 4# 4 vs 5	nd	nd	nd	2 vs 3-5	nd
PreK	nd	2-3 vs 4-5+ 3 vs 5	2 vs 3-5+# 3 vs 5 4 vs 5	nd	2-3 vs 4-5 3 vs 5	nd	nd	nd	nd	^	^	^
Toddler	nd	nd	n/a	nd	nd	n/a	^	^	n/a	^	^	n/a
Combined	nd	nd	nd	^	^	^	nd	nd	nd	^	^	^

Note. All entries (e.g. 2 vs 3-5) in this table represent statistically significant differences in CLASS scores between programs with different LD11 ratings (e.g. those rated 2 vs those rated 3, 4 or 5 for LD11), with the exception that $^{+}$ represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in the Appendix.

significant differences is in the opposite direction than expected; programs with ratings of 2 on LD11 have *higher* CLASS scores on the Organizational domain of the PreK CLASS than those with LD112 ratings of 3-5.

nd denotes no statistically significant differences.

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Table 12k4. Are there differences in CLASS scores by LD11 rating? Analysis of Variance (ANOVA) results

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	0.45	0.59	0.07	0.30	0.61	**2.26+**	**4.09***	**2.54+**	**2.44+**	1.40	1.49	0.34
N	246	246	202	120	120	107	81	81	64	45	45	35
PreK	0.42	0.10	**3.54+**	1.10	**2.57+**	1.96	1.57	0.72	**2.53+**	^	^	^
N	156	156	156	103	103	103	33	33	33			
Toddler	0.20	0.12	n/a	0.30	0.34	n/a	^	^	n/a	^	^	n/a
N	104	104		70	70							
Combined	0.37	2.29	1.78	^	^	^	0.55	0.30	0.08	^	^	^
N	63	63	63				34	34	34			

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

+ Correlation is nearing significance at the $p < .10$ level (2-tailed)

* Correlation is significant at the $p < .05$ level (2-tailed)

** Correlation is significant at the $p < .01$ level (2-tailed)

Figure 17. PreK CLASS scores by rating on LD11 (all program types)

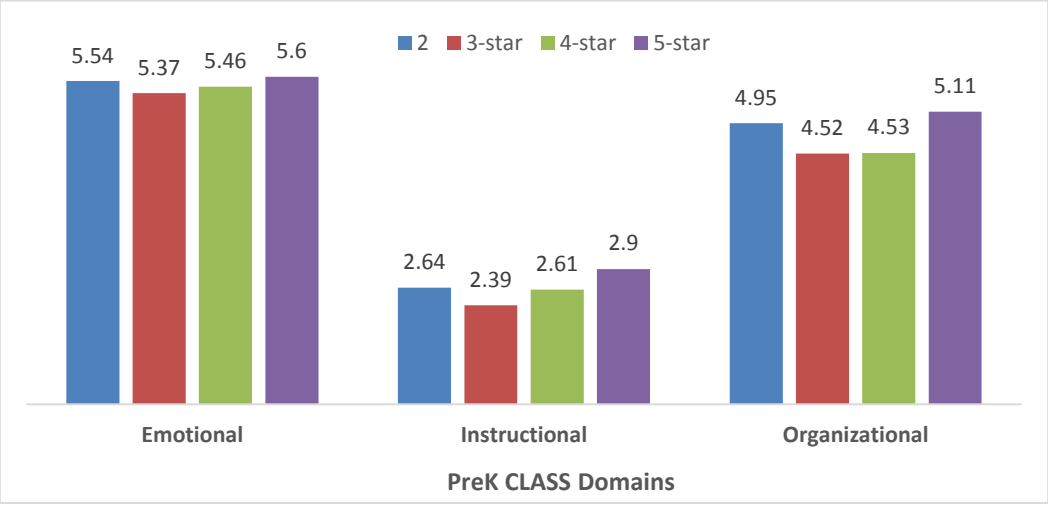


Figure 18. CLASS scores by rating on LD11 for Centers

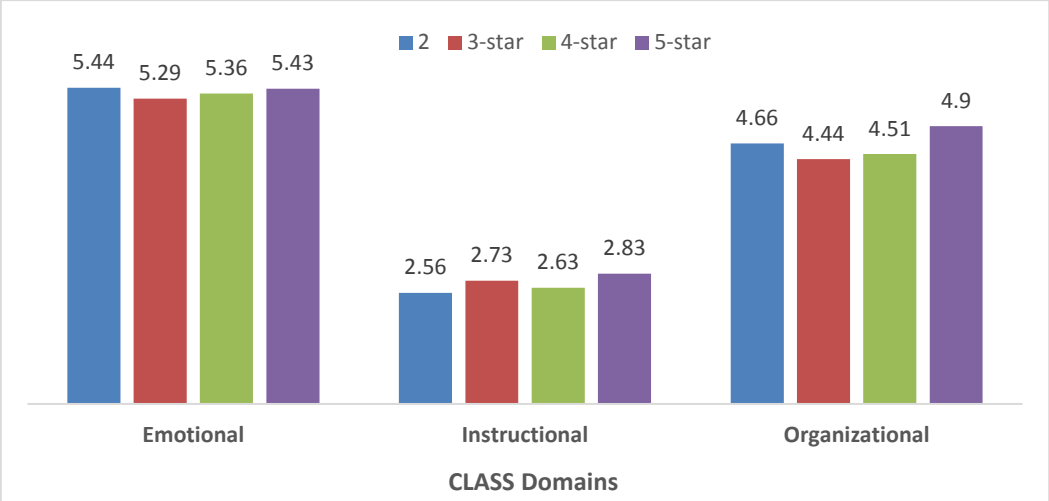


Figure 19. CLASS scores by rating on LD11 for Certified Family

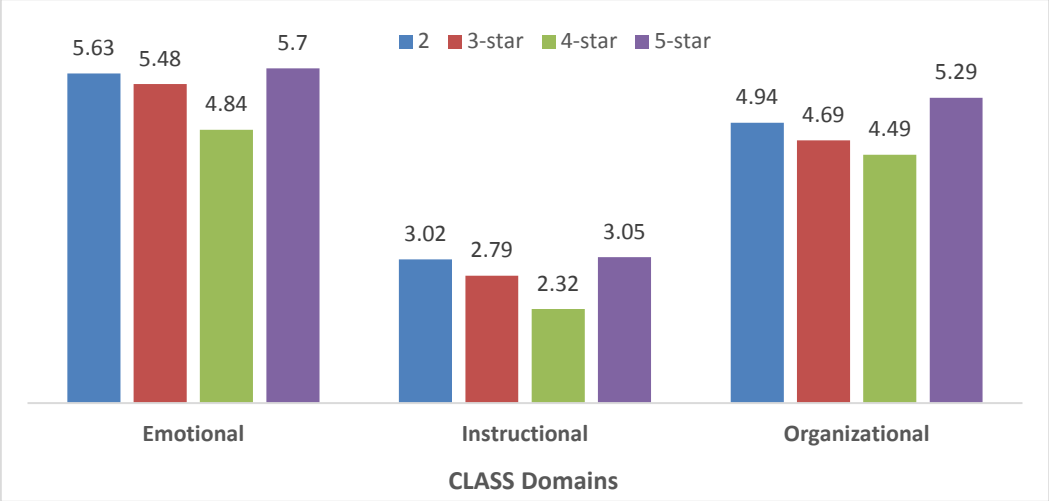
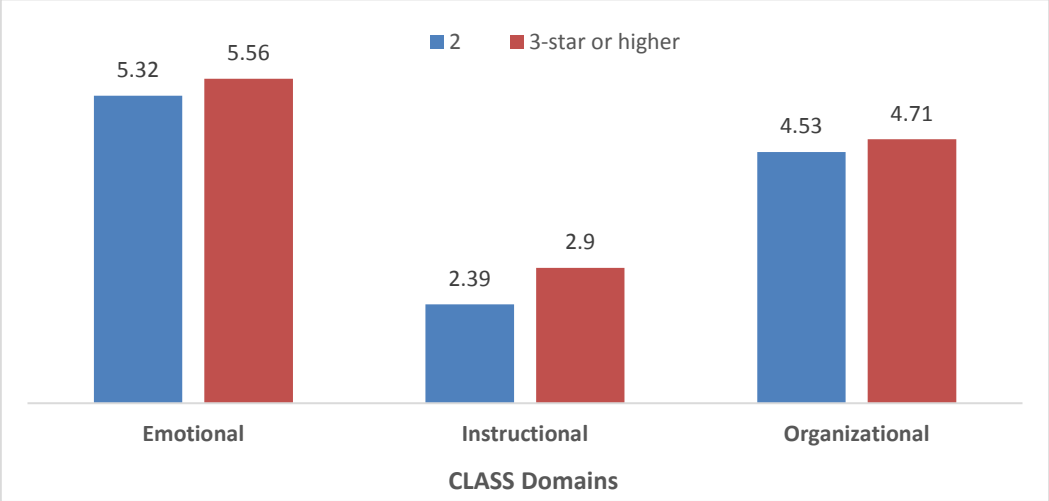


Figure 20. CLASS scores by rating on LD11 for Registered Family



Learning and Development 12: Support Social and Emotional Development

The Learning and Development 12 (LD12) standard addresses the topic of adult-child interactions that are specific to children's social and emotional development. Similar to LD11, at the 3-star Level LD12 specifies that programs must have written guidelines for behavior management that encourages the use of clear expectations, proactive and preventive strategies, and redirection. At the 4-star Level LD12 specifies that programs must directly instruct and support children in social skills, evidenced by a written description of such instruction/support. The 5-star Level involves onsite observation of the adult-child interaction using the Classroom Assessment Scoring System (CLASS), conducted by the QRIS team (not the Validation Study team). The 5-star indicators for LD12 focus on the Negative Climate and Behavioral Management dimensions of the CLASS, which are within the Emotional and Organizational domains.

Overall, 12% of programs fail to meet the 3-star Level or higher on LD12 (Table 12l1); this is substantially lower than the percent of programs that failed to meet the 3-star Level on LD11. The percent is highest in Registered Family programs (20%), followed by Certified Family (12%) and Certified Centers (8%). Overall, 8-11% of programs that are rated a 2 on LD12 have CLASS scores that are high, relative to other programs in this study (Table 12l1).

Findings revealed only a few significant differences in CLASS scores by LD12 ratings; these varied by program type and the CLASS instrument used (PreK, Toddler, or Combined). Overall, there was a small significant correlation between LD12 ratings and the Organizational domain of the CLASS (Table 12l2), which includes a behavior management dimension that is aligned conceptually with LD12 indicators. A closer look at the data (Table 12l3) reveals that most of these differences in the Organizational CLASS domain were detected in Certified and Registered Family programs (Figures 23 and 24), and with the Combined CLASS tool (Figure 22), which was primarily used in Certified and Registered Family programs. In these cases the patterns indicate that programs rated 2, and sometimes 3-star, on LD12 scored lower on the CLASS than those rated higher. For Registered Family providers, the differences in Emotional and Organizational domains shown in Table 12l3 and Figure 24 may be meaningful even though they are not quite statistically significant in this small sample of Registered Family programs. Similarly, all three of the correlations shown in Table 12l2 are similar in size (r = .18 to .27) to correlations that are statistically significant in the current study in analyses with larger samples (e.g. in Centers, Certified Family programs, or for programs overall).

For Centers, there were no statistically significant differences in CLASS scores by LD12 ratings even though the sample size (120) should have been large enough to detect significant differences. On the PreK CLASS there was also a concerning finding that programs rated 3 on LD12 showed significantly lower Instructional Support than programs rated a 2 (Table 12l2 and Figure 21). Also shown in Figure 21 is that programs rated 4-star on LD12 have similar CLASS scores to those rated 5 on LD12.

In sum, there is some indication that LD12 may differentiate CLASS scores on the Organizational domain overall, but that this primarily pertains to Certified and Registered Family programs assessed on the Combined CLASS tool for mixed age groups of children.

Table 12l1. Percentage of programs rated 2 on LD12 with high CLASS scores by program type

					Of the programs rated 2 what % had "high" quality on CLASS?											
	Percent of programs rated 2				All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD12: Support Soc. Em. Dev.	12%	20%	8%	12%	11%	8%	11%	10%	11%	11%	9%	6%	11%	14%	8%	13%

Table 12l2. Correlations among LD12 and CLASS scores by program type

	All Program Types			Certified Centers			Certified Family			Registered Family		
	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
LD12	.08	.08	**.16***	.03	.06	.14	.15	.08	**.22⁺**	.18	.22	.27
N	246	246	206	120	120	120	81	81	64	45	45	35

⁺ Correlation is nearing significance at the $p < .10$ level (2-tailed)

* Correlation is significant at the $p < .05$ level (2-tailed)

Table 12i3. Which LD12 ratings significantly differentiate CLASS scores? A summary of results come from various ANOVA tests

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS Instrument	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	2 vs 3-5+	nd	2-3 vs 4-5	nd	nd	nd	Nd	nd	2-3 vs 4-5+	2 vs 3-5+	nd	2 vs 3-5+
PreK	nd	2-3 vs 4-5# 3 vs 5+	nd	2 vs 3-5+	2-3 vs 4-5+	nd	Nd	2 vs 3-5#	nd	^	^	^
Toddler	nd	nd	n/a	nd	nd	n/a	^	^	n/a	^	^	n/a
Combined	nd	2 vs 3-5+	2 vs 3-5 2-3 vs 4-5+	^	^	^	Nd	nd	nd	^	^	^

Note. All entries (e.g. 2 vs 3-5) in this table represent statistically significant differences in CLASS scores between programs with different LD12 ratings (e.g. those rated 2 vs those rated 3, 4 or 5 for LD12), with the exception that ⁺ represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in the Appendix.

significant differences is in the opposite direction than expected

nd denotes no statistically significant differences.

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Table 12i4. Are there differences in CLASS scores by LD12 rating? Analysis of Variance (ANOVA) results

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	1.22	0.68	1.93	0.56	0.52	0.70	1.49	0.61	1.06	**3.94+**	1.01	0.83
N	246	246	202	120	120	107	81	81	64	45	45	35
PreK	1.00	**2.78***	0.67	1.70	**2.15+**	0.32	1.30	1.78	1.43	^	^	^
N	156	156	156	103	103	103	33	33	33			
Toddler	0.81	0.79	n/a	0.76	1.68	n/a	^	^	n/a	^	^	n/a
N	104	104		70	70							
Combined	1.22	1.15	1.75	^	^	^	0.13	0.49	0.47	^	^	^
N	63	63	63				34	34	34			

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

+ Correlation is nearing significance at the $p < .10$ level (2-tailed)

* Correlation is significant at the $p < .05$ level (2-tailed)

** Correlation is significant at the $p < .01$ level (2-tailed)

Figure 21. PreK CLASS scores by rating on LD12 (all program types)

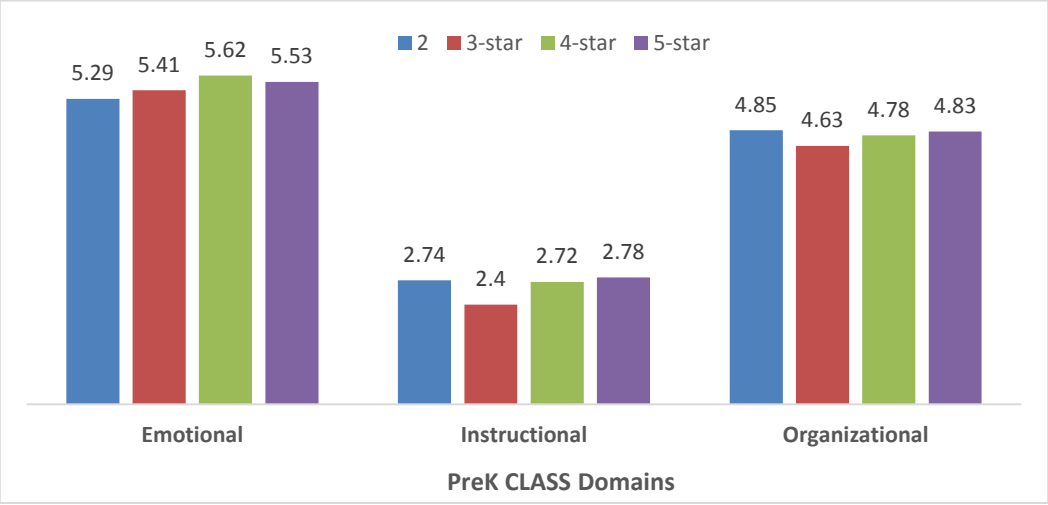


Figure 22. Combined CLASS scores by rating on LD12 (all program types)

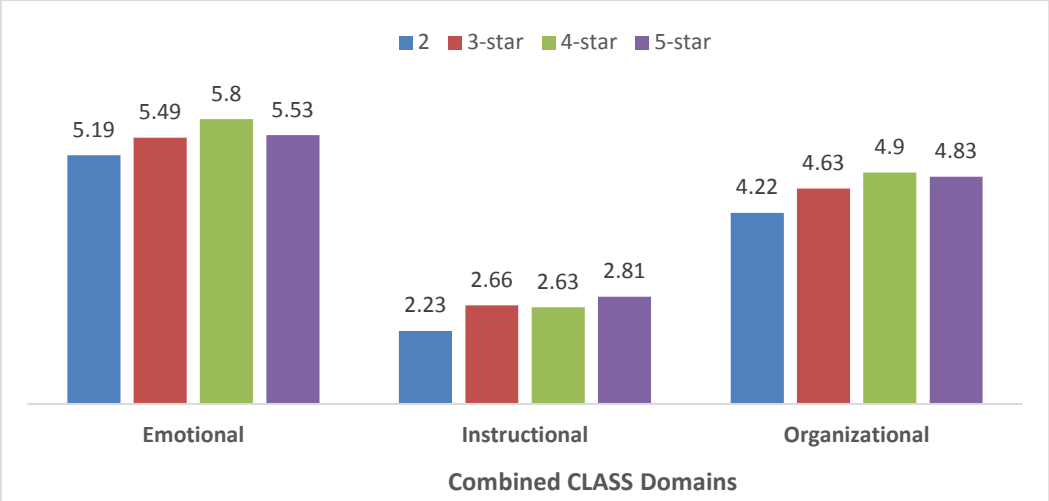
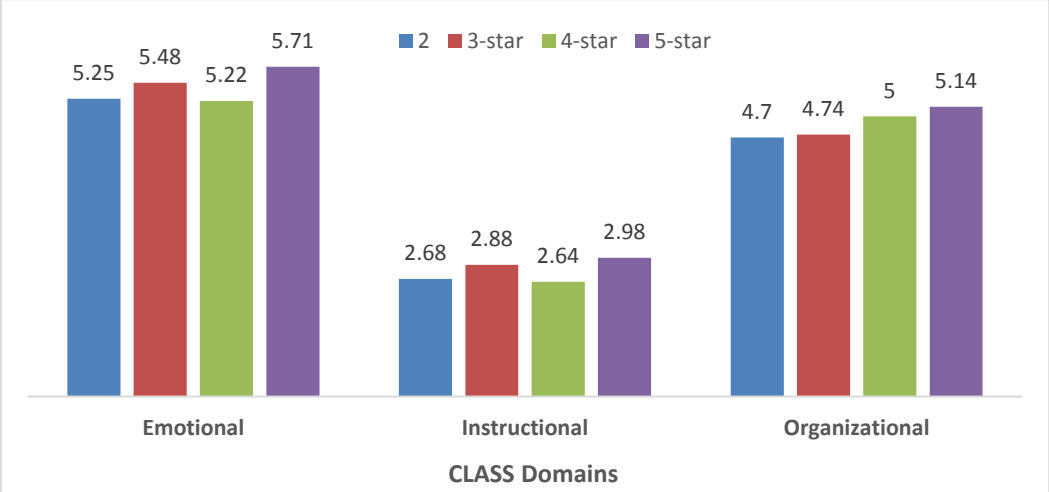
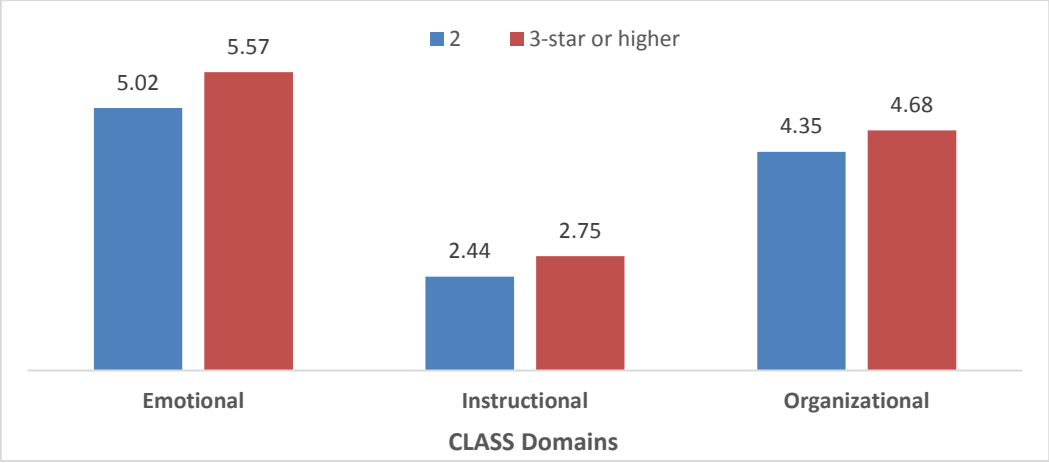


Figure 23. CLASS scores by rating on LD12 for Certified Family



Note. CLASS scores are the average of the PreK, Toddler, and Combined instruments

Figure 24. CLASS scores by rating on LD12 for Registered Family



Note. CLASS scores are the average of the PreK, Toddler, and Combined instruments

Personnel qualifications

Personnel Qualifications 1: Leader Qualifications

The percent of programs rated a 2 on PQ1 (Table 12m1) is 16% overall, but is high (27%) for Registered Family providers. For Centers and Certified Family providers relatively few programs rated 2 received high scores on the CLASS, relative to other programs in this study. This indicates good alignment between the PQ1 ratings and CLASS scores. The Correlations presented in Table 23b and the ANOVAs presented in Table 12m2 show statistically significant, although modest (r = .17 to r = .25), associations between ratings on PQ1 and CLASS scores, such that higher PQ1 ratings are associated with higher CLASS scores. The evidence is stronger for Centers than for Certified Family providers. Further, it appears that the largest differences in CLASS scores are between programs rated a 2 vs 3-star or higher on PQ1 (Table 12m3 and Figure 25). Nearly all of the links between PQ1 and the CLASS are for the PreK version of the CLASS, and for Centers. The only statistically significant difference in CLASS scores for Certified Family providers is that programs with ratings of 2 or 3 on PQ1 had significantly lower scores on the Organizational Support domain of the CLASS.

These results suggest that a primary concern about PQ1 is that it does not seem to work as well for Registered Family providers. Twenty-seven percent of Registered Family providers were rated a 2 on this standard (Table 12m1). Evidence from the QRIS process evaluation indicates that 40% of Registered Family providers pass all other QRIS domains at the 3-star Level or higher but that they end up with a 2 overall on the QRIS because they were rated a 2 on PQ1 (Guy & Aldrich, April, 2015). This pattern is most concerning because a sizeable proportion of these Registered Family programs that rated a 2 on PQ1 received relatively high scores on the CLASS. In other words, there appears to be a mismatch between PQ1 ratings and CLASS scores for many of the Registered Family providers. This mismatch is also evidenced by a lack of correlation, and a lack of significant results from ANOVA tests for Registered Family providers (Tables 12m2 and 12m3).

In sum, there is evidence for small, significant links between PQ1 and CLASS scores for Centers and Certified Family programs.

Table 12m1. Percentage of programs rated 2 on PQ1 with high CLASS scores, by program type

					Of the programs rated 2 what % had "high" quality on CLASS?											
	Percent of programs rated 2				All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
PQ1	16%	27%	16%	11%	14%	9%	11%	15%	11%	11%	5%	5%	7%	29%	23%	17%

Table 12m2. Correlations among PQ1 and CLASS scores by program type

	All Program Types			Certified Centers			Certified Family			Registered Family		
	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
PQ1	.10	**.17****	**.25****	**.15**[+]**	.13	**.23***	.16	**.21**[+]**	.35	-.07	.25	-.19
N	246	246	206	120	120	120	81	81	64	45	45	35

[+] Correlation is nearing significance at the $p < .10$ level (2-tailed)

** Correlation is significant at the $p < .01$ level (2-tailed)

Table 12m3.Which PQ1 ratings significantly differentiate CLASS scores? A summary of results come from various ANOVA tests

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	2-3 vs 4-5[+]	2-3 vs 4-5 2 vs 3-5 2 vs 5	2-3 vs 4-5 2 vs 3-5 2 vs 3[+] 2 vs 4 2 vs 5	2-3 vs 4-5 2 vs 3-5	2 vs 3-5	2-3 vs 4-5 2 vs 3-5 2 vs 5	nd	2 vs 3-5[+]	2-3 vs 4-5 2 vs 3-5[+]	nd	nd	nd
PreK	2-3 vs 4-5 2 vs 3-5	2-3 vs 4-5 2 vs 3-5 2 vs 4 2 vs 5 3 vs 5[+]	2-3 vs 4-5 2 vs 3-5 2 vs 4 2 vs 5	2-3 vs 4-5 2 vs 3-5	2-3 vs 4-5 2 vs 3-5 2 vs 5	2 vs 3-5	nd	nd	nd	^	^	^
Toddler	nd	nd		nd	nd	nd	^	^	^	^	^	^
Combined	nd	nd	2-3 vs 4-5[+]	^	^	^	nd	nd	nd	^	^	^

Note. All entries (e.g. 2-3 vs 4-5) in this table represent statistically significant differences in CLASS scores between programs with different PQ1 ratings (e.g. those rated 2 or 3 vs those rated 4 or 5 for PQ1), with the exception that [+] represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in the Appendix.

nd denotes no statistically significant differences.

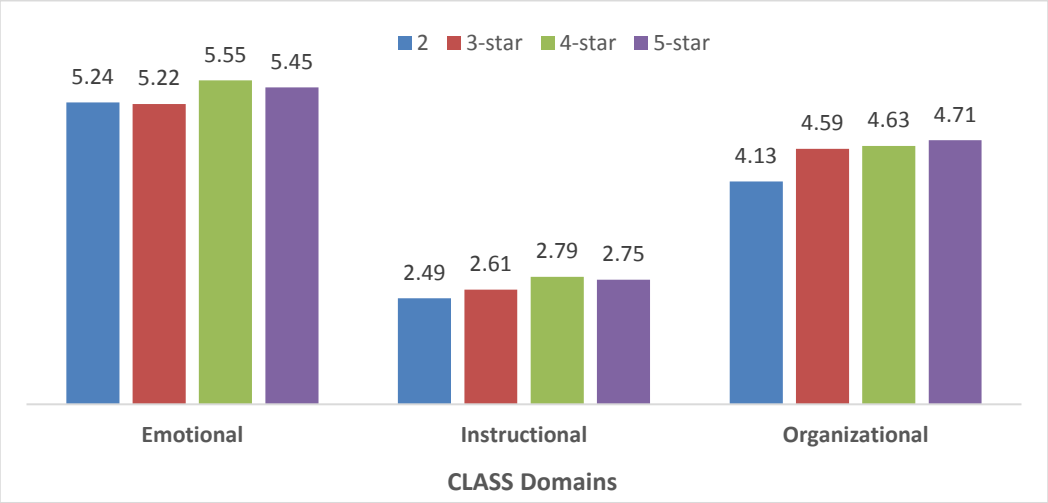
^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Table 12m4. Are there differences in CLASS scores by PQ1 rating? Analysis of Variance (ANOVA) results

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	**CLASS domains**			**CLASS domains**			**CLASS domains**			**CLASS domains**		
Instrument	**ES**	**IS**	**OC**	**ES**	**IS**	**OC**	**ES**	**IS**	**OC**	**ES**	**IS**	**OC**
Average	1.04	**2.77***	**5.89****	1.40	0.83	**2.79***	0.85	1.48	**3.04***	0.47	0.93	1.19
N	246	246	202	120	120	107	81	81	64	45	45	35
PreK	**2.72***	**5.41***	**3.60***	1.58	**3.51****	1.54	**3.90***	1.94	**2.39⁺**	^	^	^
N	156	156	156	103	103	103	33	33	33			
Toddler	0.10	0.59	n/a	0.66	0.22	n/a	^	^	n/a	^	^	n/a
N	104	104		70	70							
Combined	0.44	0.84	1.45	^	^	^	0.77	0.90	0.73	^	^	^
N	63	63	63				34	34	34			

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).
⁺ Correlation is nearing significance at the $p < .10$ level (2-tailed)
* Correlation is significant at the $p < .05$ level (2-tailed)
** Correlation is significant at the $p < .01$ level (2-tailed)

Figure 25. CLASS Scores by Rating on PQ1 (Director Qualifications) in Centers



Note. CLASS scores are the average of the PreK, Toddler, and Combined instruments

Personnel Qualifications 2: Teacher Qualifications

The PQ2 standard regarding teacher qualifications is only applicable to Centers. Although a few Certified and/or Registered Family providers did receive ratings for PQ2 the Validation study did not consider these ratings because "teacher" is not an allowable position title for licensed (certified or registered) family providers, and because the QRIS scoring guide indicates that PQ2 is not applicable to family providers.

The percent of Centers rated a 2 on PQ2 (Table 12n1) is a bit high (18%) compared to other standards (Figure 1). However, fewer than 15% of the programs rated a 2 scored high on the Instructional and Organizational domains of the CLASS, relative to other programs in this study. Thirty percent of them did score high on Emotional Support, but this is less concerning because Emotional Support scores were fairly high among programs participating in this study overall.

PQ2 ratings are not significantly correlated with programs' CLASS scores (Table 12n2) but the ANOVA tests did uncover some differences in programs' CLASS scores by PQ2 ratings (Table 12n3). The pattern of findings (Table 12n3) indicate that programs rated 2 on PQ2 have lower CLASS scores than those rated 3 or higher. This is primarily the case for the Instructional domain on the PreK CLASS, but there is also a hint of evidence for the Organizational (average across all CLASS instruments) and Emotional (PreK CLASS only; not quite statistically significant) domains. This pattern of findings is illustrated for Centers in Figure 26.

A challenge in detecting links between PQ2 and CLASS scores is that PQ2 requires 50% of teachers in a program to have certain qualifications (e.g. a step 7.5 or higher on the Registry) and the CLASS scores were obtained through observation of randomly selected classrooms. Thus, the specific staff included in PQ2 ratings may not have been the same as those observed with the CLASS.

In sum, there is some limited evidence for small significant links between PQ2 and CLASS scores in Centers.

Table 12n1. Percentage of programs rated 2 on PQ2 with high CLASS scores by program type

					Of the programs rated 2 what % had "high" quality on CLASS?											
	Percent of programs rated 2				All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
PQ2 Teachers	n/a	n/a	18%	n/a	n/a	n/a	n/a	30%	14%	13%	n/a			n/a		

n/a refers to a standard that is not applicable. PQ2 does not apply for Family Child Care Providers.

Table 12n2. Correlations among PQ2 and CLASS scores by program type

	All Program Types			Certified Centers			Certified Family			Registered Families		
	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
PQ2	n/a			.03	.13	.13	n/a			n/a		
N				119	119	106						

n/a refers to a standard that is not applicable. PQ2 does not apply for Certified and Registered Family Programs.

Table 12n3.Which PQ2 ratings significantly differentiate CLASS scores? A summary of results come from various ANOVA tests

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	n/a			nd	2 vs 3-5^{+}	2 vs 3-5	n/a			n/a		
PreK				2 vs 3-5^{+}	2-3 vs 4-5 2 vs 3-5 2 vs 5	2 vs 3-5^{+}						
Toddler				nd	nd	nd						
Combined				^	^	^						

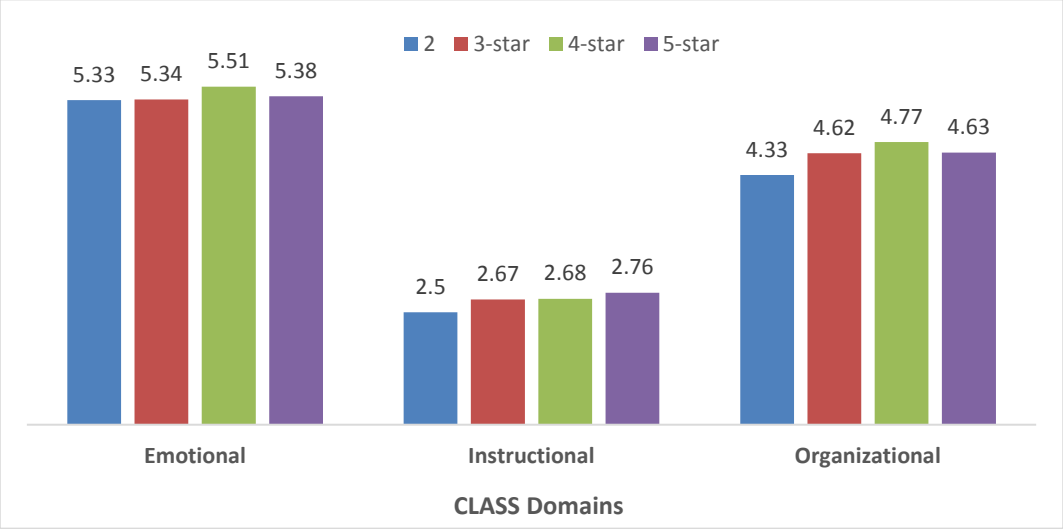
Note. All entries (e.g. 2-3 vs 4-5) in this table represent statistically significant differences in CLASS scores between programs with different PQ2 ratings (e.g. those rated 2 or 3 vs those rated 4 or 5), with the exception that $^{+}$ represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in the Appendix.
nd denotes no statistically significant differences.
^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).
n/a refers to a standard that is not applicable. PQ2 does not apply for Family Child Care Providers.

Table 12n4. Are there differences in CLASS scores by PQ2 rating? Analysis of Variance (ANOVA) results

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	0.23	1.01	1.90	0.23	0.70	1.30	n/a	n/a	n/a	n/a	n/a	n/a
N	160	160	139	119	119	106						
PreK	1.14	**3.59***	1.22	1.40	**3.12***	0.87	n/a	n/a	n/a	n/a	n/a	n/a
N	118	118	118	103	103	103						
Toddler	0.17	0.42	n/a	0.04	0.52	n/a	n/a	n/a	n/a	n/a	n/a	n/a
N	78	78		70	70							
Combined	0.09	0.26	1.33	^	^	^	n/a	n/a	n/a	n/a	n/a	n/a
N	29	29	29									

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).
+ Correlation is nearing significance at the $p < .10$ level (2-tailed)
* Correlation is significant at the $p < .05$ level (2-tailed)
** Correlation is significant at the $p < .01$ level (2-tailed)

Figure 26. CLASS Scores by rating on PQ2 (Teacher Qualifications) in Centers



Note. CLASS scores are the average of the PreK, Toddler, and Combined

Personnel Qualifications 3: Aide/Assistant Qualifications

The PQ3 standard regarding qualifications of assistants is primarily applicable to Centers and Certified Family providers. Only those programs that use aides/assistants are rated on PQ3; the others receive a rating of N/A and are therefore excluded from the analyses for this standard.

The percent of Centers rated a 2 on PQ3 (Table 12o1) is fairly small compared with other standards. Few of these programs rated a 2 scored high on the CLASS, relative to other programs in this study.

PQ3 ratings are significantly correlated with programs' CLASS scores on the Organizational domain (Table 12o2); programs with higher PQ3 scores also tended to be rated more highly on the CLASS Organizational domain. This correlation appears to be primarily driven by the Certified Family programs, where these differences were strongest, although the ANOVA results also reveal some differences in CLASS scores by PQ3 rating for Centers (Table 12o3 and Figure 27). There were differences on all three domains of the PreK CLASS score by PQ3 ratings in both Centers and Certified Family programs, although the differences for Emotional Support in Centers and for Instructional Support in Certified Family were not quite statistically significant.

In sum, there is some evidence for small, significant links between PQ3 and CLASS scores in Centers and Certified Family providers.

Table 12o1. Percentage of programs rated 2 on PQ3 with high CLASS scores by program type

					Of the programs rated 2 what % had "high" quality on CLASS?											
	Percent of programs rated 2				All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
PQ3-Aides	12%	n/a	10%	6%	3%	6%	4%	9%	11%	4%	0%	4%	6%	n/a		

n/a refers to a standard that is not applicable.

Table 12o2. Correlations among PQ3 and CLASS scores by program type

	All Program Types			Certified Centers			Certified Family			Registered Families		
	CLASS domains			CLASS domains			CLASS domains			CLASS Domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
PQ3	.12	.15	**.22****	.11	.11	.10	.16	.16	**.39****	^	^	^
N	153	153	133	81	81	74	59	59	48			

** Correlation is significant at the $p < .01$ level (2-tailed)

^ indicates too few programs to support analysis.

Table 12o3.Which PQ3 ratings significantly differentiate CLASS scores? A summary of results come from various ANOVA tests

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS Instrument	CLASS domains ES	IS	OC	CLASS domains ES	IS	OC	CLASS domains ES	IS	OC	CLASS domains ES	IS	OC
Average	nd	nd	2 vs 3-5 2 vs 5+	2 vs 3-5	nd	2-3 vs 4-5 2 vs 3-5 2 vs 5	nd	nd	2-3 vs 4-5 2 vs 5	^	^	^
PreK	2 vs 5+	2 vs 5+	2 vs 3-5 2 vs 5	2 vs 5+ 2 vs 3-5+	2-3 vs 4-5 2 vs 3-5+ 3 vs 5	2 vs 5+ 2 vs 3-5	2 vs 3 2 vs 5	2 vs 5+	2 vs 3 2 vs 5	^	^	^
Toddler	nd	nd	nd	nd	nd	nd	^	^	^	^	^	^
Combined	nd	nd	nd	nd	nd	nd	^	^	^	^	^	^

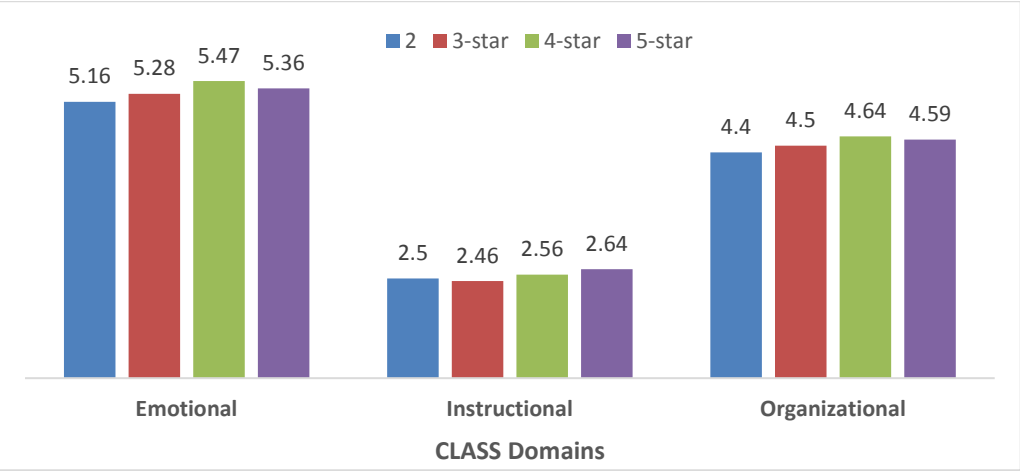
Note. All entries (e.g. 2-3 vs 4-5) in this table represent statistically significant differences in CLASS scores between programs with different PQ3 ratings (e.g. those rated 2 or 3 vs those rated 4 or 5), with the exception that + represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in the Appendix.
nd denotes no statistically significant differences.
^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Table 12o4. Are there differences in CLASS scores by PQ3 rating? Analysis of Variance (ANOVA) results

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS Instrument	CLASS domains ES	IS	OC	CLASS domains ES	IS	OC	CLASS domains ES	IS	OC	CLASS domains ES	IS	OC
Average	1.42	1.32	**2.72***	0.52	0.40	0.34	1.78	1.39	**3.45***	^	^	^
N	153	153	133	81	81	74	59	59	48			
PreK	**2.47+**	**2.77***	**2.62+**	0.95	**2.69+**	0.26	**7.96****	**2.95+**	**7.68****	^	^	^
N	104	104	104	71	71	71	27	27	27			
Toddler	0.61	0.05	n/a	0.74	0.13	n/a	^	^	n/a	^	^	n/a
N	68	68		51	51							
Combined	0.58	0.30	1.51	^	^	^	^	^	^	^	^	^
N	38	38	38									

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).
+ Correlation is nearing significance at the $p < .10$ level (2-tailed)
* Correlation is significant at the $p < .05$ level (2-tailed)
** Correlation is significant at the $p < .01$ level (2-tailed)

Figure 27. CLASS scores by rating on PQ3 (AIde Qualifications) in Centers



Note. CLASS scores are the average of the PreK, Toddler, and Combined

Personnel Qualifications 4: Personnel Training

The PQ4 standard applies to all three types of programs. The percent of programs rated a 2 on PQ4 (Table 12p1) is fairly low, and few of these programs rated a 2 scored high on the CLASS, with the exception of Emotional Support for Centers, which tends to be high overall in this sample.

PQ4 ratings are significantly correlated with programs' CLASS scores on the Organizational Support domain (Table 12p2), and their link with the Instructional Support domain of the CLASS are nearing significance for Certified Family providers and the average across all program types. Thus programs with higher PQ4 ratings tended to be rated higher on the CLASS especially in the Organizational Support domain. Although correlations are not significant for Registered Family providers they are roughly similar in size for Registered Family; detecting statistical significance with a sample of 35 to 45 programs is challenging. ANOVA results (Table 12p3) also reveal differences in all three domains of the PreK CLASS by PQ4 rating for Centers, as well as for the Organizational Support domain of the PreK CLASS for Certified Family providers, and the Emotional Support domain for Registered Family providers (averaged across CLASS instruments). Figure 28 illustrates these patterns for the PreK CLASS. Thus, there is some preliminary evidence that PQ4 may meaningfully distinguish CLASS ratings for Registered Family providers as well.

In sum, there is some evidence for small, significant links between PQ4 and CLASS scores across all types of programs. Evidence is most consistent for PreK classrooms in Centers, and for the Organizational domain in PreK-aged groups in Centers and Certified Family programs.

Table 12p1. Percentage of programs rated 2 with high CLASS in PQ4 (training) by program type

	Percent of programs rated 2				Of the programs rated 2 what % had "high" quality on CLASS?											
					All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
PQ4	11%	9%	13%	10%	9%	8%	7%	20%	14%	11%	5%	6%	7%	0%	0%	0%

Table 12p2. Correlations among PQ4 and CLASS scores by program type

	All Program Types			Certified Centers			Certified Family			Registered Family		
	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
PQ4	.07	**.12⁺**	**.22****	-.06	.07	**.20***	.11	**.19⁺**	**.32***	.08	.19	.18
N	246	246	206	120	120	120	81	81	64	45	45	35

⁺ Correlation is nearing significance at the $p < .10$ level (2-tailed)

* Correlation is significant at the $p < .05$ level (2-tailed)

Table 12p3. Which PQ4 ratings significantly differentiate CLASS scores? A summary of results come from various ANOVA tests

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
Instrument	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	nd	Nd	2 vs 4 2 vs 5	Nd	2-3 vs 4-5	2-3 vs 4-5 2 vs 3-5 2 vs 4 2 vs 5	nd	2-3 vs 4-5+	2-3 vs 4-5	2 vs 3 2 vs 4	nd	nd
PreK	2 vs 4	2 vs 4+ 2 vs 5+	2 vs 4 2 vs 5+	2-3 vs 4-5 2 vs 3-5 2 vs 4 2 vs 5	2-3 vs 4-5 2 vs 3-5 2 vs 4 2 vs 5	2-3 vs 4-5 2 vs 3-5 2 vs 4 2 vs 5	nd	nd	2-3 vs 4-5	^	^	^
Toddler	nd	nd	nd	Nd	Nd	nd	^	^	^	^	^	^
Combined	nd	nd	nd	^	^	^	nd	nd	nd	^	^	^

Note. All entries (e.g. 2-3 vs 4-5) in this table represent statistically significant differences in CLASS scores between programs with different PQ4 ratings (e.g. those rated 2 or 3 vs those rated 4 or 5), with the exception that ⁺ represents differences that are *nearing* significance at the $p < .10$ level (2-tailed). All others listed are statistically significant at the $p < .05$ or $p < .01$ level. Specific ANOVA results are available in the Appendix.

nd denotes no statistically significant differences.

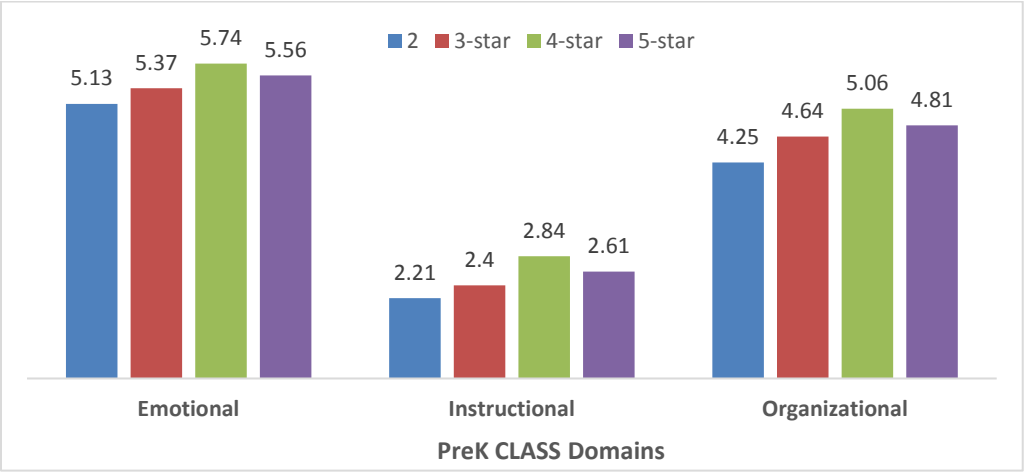
^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).

Table 12p4. Are there differences in CLASS scores by PQ4 rating? Analysis of Variance (ANOVA) results

	All Program Types			Certified Centers			Certified Family			Registered Family		
CLASS Instrument	CLASS domains			CLASS domains			CLASS domains			CLASS domains		
	ES	IS	OC	ES	IS	OC	ES	IS	OC	ES	IS	OC
Average	1.24	**2.54+**	**5.11***	1.76	**2.27+**	**2.94***	0.43	0.96	**2.45+**	**5.19***	**2.31+**	2.07
N	246	246	206	120	120	107	81	81	64	45	45	35
PreK	**3.45***	**4.42***	**3.52***	**3.94***	**4.82***	**2.86***	0.87	0.97	2.22	^	^	^
N	156	156	156	103	103	103	33	33	33			
Toddler	0.53	1.81	n/a	1.07	1.77	n/a	^	^	n/a	^	^	n/a
N	104	104		70	70							
Combined	0.59	0.67	1.25	^	^	^	0.13	0.53	0.22	^	^	^
N	63	63	63				34	34	34			

^ indicates comparisons with too few programs to conduct ANOVA tests (N < 25).
+ Correlation is nearing significance at the $p < .10$ level (2-tailed)
* Correlation is significant at the $p < .05$ level (2-tailed)
** Correlation is significant at the $p < .01$ level (2-tailed)

Figure 28. PreK CLASS scores by rating on PQ4 (all program types)



Personnel Qualifications 5: Personnel training in ethics, professional responsibility, and confidentiality

The PQ5 standard applies to all three types of programs. The percent of programs rated a 2 on PQ5 (Table 12q1) is low (9%) compared with other standards, and very few of these programs rated a 2 scored high on the CLASS. The Validation Study does not report associations between programs PQ5 ratings and their CLASS scores because the PQ5 standard, and the evidence programs must submit to achieve it, does not have a strong theoretical link with the quality of adult-child interactions.

Table 12q1. Percentage of programs rated 2 on LD5 with high CLASS scores, by program type

	Percent of programs rated 2				Of the programs rated 2 what % had "high" quality on CLASS?											
					All Programs			Certified Centers			Certified Family			Registered Family		
	Total	RF	CC	CF	ES	IS	OS	ES	IS	OS	ES	IS	OS	ES	IS	OS
PQ 5	9%	2%	10%	10%	4%	2%	4%	5%	5%	4%	5%	0%	4%	0%	0%	0%

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

Table 13a. Correlations within QRIS Personnel Quality Measures

	PQ1	PQ2	PQ3	PQ4
PQ1	**1**			
PQ2	**0.64*** **	**1**		
PQ3	**0.71*** **	**0.62*** **	**1**	
PQ4	**0.66*** **	**0.66*** **	**0.60*** **	**1**

N-value varies for each correlation based on how many were rated.: Minimum N=120 and Maximum N=246.
*** Correlation is significant at the p<0.001 level (2-tailed)

Table 13b. Correlations within VS-ORO Personnel Quality Measures

	Dir/ Provider Step	Teacher Median Step	Aide/ Assistant Median Step	% Staff with 24 hours or more of training
VS-ORO Director/Provider Step	**1**			
VS-ORO Teacher Median Step	**0.46*** **	**1**		
VS-ORO Aide/Assistant Median Step	0.14	0.10	**1**	
% Staff with 24 hours or More of Training	-0.02	0.05	**0.25** **	**1**

N-value varies for each correlation based on how many were rated.: Minimum N=72 and Maximum N=246.
** Correlation is significant at the p<0.01 level (2-tailed)
*** Correlation is significant at the p<0.001 level (2-tailed)

Table 13c1. Correlations within Structural Indicators Personnel Quality Measures, Centers

	Dir has Step 9 or Higher	Teachers have Step 9 or Higher	Aides have a Step 5 or Higher	% Staff with 20 Hours or More Training
Director has Step 9 or Higher	**1**			
% Teachers with Step 9 or Higher	**0.36*** **	**1**		
% Aides with Step 5 or Higher	0.14	0.13	**1**	
% Staff with 20 hours or More of Training	**0.36*** **	**0.46*** **	**0.35** **	**1**

N-value varies for each correlation based on how many were rated.: Minimum N=68 and Maximum N=140.
** Correlation is significant at the p<0.01 level (2-tailed)
*** Correlation is significant at the p<0.001 level (2-tailed)

Table 13c2. Correlations within Structural Indicators Personnel Quality Measures, Certified Family

	Provider Step 9 or Higher	Assistant Step 5 or Higher	% Staff 20 Hours or More Training
Provider has Step 9 or Higher	**1**		
Assistant has Step 5 or Higher	**0.46*** **	**1**	
% Staff with 20 Hours or More of Training	0.12	**0.26***	**1**

N-value varies for each correlation based on how many were rated.: Minimum N=56 and Maximum N=77.
* Correlation is significant at the p<0.05 level (2-tailed)
*** Correlation is significant at the p<0.001 level (2-tailed)

Table 13c3. Correlations within Structural Indicators Personnel Quality Measures, Registered Family

Correlation	Provider Step 8 or Higher	Provider has 20 Hours or More Training
Provider has Step 8 or Higher	**1**	
Provider has 20 Hours or More of Training	0.17	**1**

N-value varies for each correlation based on how many were rated.: N=58.

Description of Personnel Measures by PQ Construct

***PQ1**:* **Measures of director/provider qualifications across the three datasets.** Center director qualification measures varied across the three data sets. For QRIS, a PQ1 rating of 3 was based on directors having a step of 8 or higher whereas for Family providers it was having a step 7 or higher. A PQ1 rating of 4 was based on having a step 9 or higher for Centers and an 8 or higher for Family and a rating of 5 was based on having a step 10 or higher for Centers and 9 or higher for Family. The VS-ORO measure was of the director/provider's actual step. The SI dataset included three measures related to steps (percent of Center directors with a step 8 or higher, those with a step 9 or higher, or those with a step 10 or higher). In keeping with QRIS ratings for family providers, SI measures for family providers were lower (percent of CF/RF providers with a step 7 or higher, those with a step 8 or higher, those with a step 9 or higher). SI measures included percent of directors since some centers had multiple directors.

***PQ2**:* **Measures of teacher qualification across the three datasets.** The QRIS measure for Center teachers for a QRIS PQ2 rating of 3 was based on the percentage of teachers having a step 7 or higher, with having a step 8 required for a rating of 4, and having a step 9 or higher for a rating of 5. VS-ORO measures included the average and median step of all teachers in the facility whereas SI included three measures related to steps (percent of teachers with a step 7 or higher, those with a step 8 or higher, or those with a step 9 or higher).

***PQ3**:* **Measures of Center aide and Certified Family assistant qualifications across the three datasets.**

The PQ measures for Center aide and Certified Family assistant qualifications for a QRIS PQ rating of 3 were based on the percent having a step 3 or higher. VS-ORO measures included the average and median step of all aides or assistants in the facility whereas SI included three measures related to steps (percent of aides/assistants with a step 3 or higher, those with a step 5 or higher, or those with a step 7or higher).

***PQ4:* Measures of training across the three datasets.**
QRIS PQ4, VS-ORO, and SI contained measures of training but there were key differences in the measures used. The QRIS training standard (PQ4) for centers and homes required that 75% of staff (directors, teachers, aides, assistants) have a set number of hours of training (community-based or college hours): 18 for a rating of 3, 20 for a rating of 4, and 24 for a rating of 5. Three VS-ORO training measures captured the percent of staff in centers and homes with a set number of training hours: 18, 20, and 24. Similarly, SI captured the percent of center staff (directors, teachers, aides) and family home providers and assistants that had a set number of training hours: 18, 20, and 24. It is important to note that the SI measure included both community-based hours and those earned in college courses whereas the other two data sources counted only community-based hours.

Correlations Among Measures of Director/Provider Qualifications

Following are details from the analysis of the correlations among the three sets of personnel quality. They are organized in two ways: a) by the PQ construct (i.e., PQ1, PQ2), and b) by correlations between two sets of measures (i.e., QRIS PQ1 with VS-ORO, QRIS PQ1 with SI, and VS-ORO with SI).

PQ1: Director/Provider Qualifications

Table 13d. Correlations between QRIS and VS-ORO Measures of Director/Provider Qualifications

Correlation	Center Director	CF Provider	RF Provider
QRIS PQ1 with VS-ORO Provider Step	**0.45*** **	**0.73*** **	**0.73*** **

N-value varies for each correlation based on how many were observed. For Centers: N=91. For CF: Minimum N=74. For RF: Minimum N=43.
*** Correlation is significant at the p<0.001 level (2-tailed)

Table 13e. Correlations between QRIS and SI Measures of Director/Provider Qualifications

Correlation	Center Director	CF Provider	RF Provider
QRIS PQ1 with SI Provider Step 7	--	**0.54*** **	**0.53*** **
QRIS PQ1 with SI Provider Step 8	**0.35***	**0.71*** **	**0.71*** **
QRIS PQ1 with SI Provider Step 9	**0.40***	**0.75*** **	**0.72*** **
QRIS PQ1 with SI Provider Step 10	**0.30***	--	--
QRIS PQ1 with Director/Provider with Some College or Degree	0.09	**0.47*** **	0.26^{+}
QRIS PQ1 with Director/Provider Degree	**0.33*** **	**0.37** **	0.24
QRIS PQ1 with Director/Provider ECE–Related Degree	**0.29** **	**0.39** **	**0.37***

N-value varies for each correlation based on how many were observed. For Centers: Minimum N=109 and Maximum N=112. For CF: N=68. For RF: Minimum N=43 and Maximum N=44.
$^{+}$ Correlation is nearing significance at the p<0.10 level (2-tailed)
* Correlation is significant at the p<0.05 level (2-tailed)
** Correlation is significant at the p<0.01 level (2-tailed)
*** Correlation is significant at the p<0.001 level (2-tailed)

Table 13f. Correlations between VS_ORO and SI Measures of Director/Provider Qualifications.

Correlation	Center Director	CF Provider	RF Provider
VS-ORO Director/Provider Step with SI Provider Step 7	--	**0.64*** **	**0.55*** **
VS-ORO Director/Provider Step with SI Provider Step 8	**0.58***	**0.70*** **	**0.62*** **
VS-ORO Director/Provider Step with SI Provider Step 9	**0.44***	**0.69*** **	**0.53*** **
VS-ORO Director/Provider Step with SI Provider Step 10	**0.38***	--	--
VS-ORO Director/Provider Step Dir/Provider w Some College or Degree	0.05	**0.58*** **	0.27+
VS-ORO Director Step with Director Provider Degree	**0.38*** **	**0.64*** **	**0.45** **
VS-ORO Director Step with Director/Provider ECE-Related Degree	**0.42*** **	**0.48*** **	**0.60*** **

N-value varies for each correlation based on how many were observed. For Centers: Minimum N=91 and Maximum N=140. For CF: Minimum N=57 and Maximum N=81. For RF: Minimum N=33 and Maximum N=58.

+ Correlation is nearing significance at the p<0.10 level (2-tailed)

* Correlation is significant at the p<0.05 level (2-tailed)

** Correlation is significant at the p<0.01 level (2-tailed)

*** Correlation is significant at the p<0.001 level (2-tailed)

PQ2: Center Teacher Qualifications

Table 13g. Correlations between QRIS and VS-ORO Measures of Center Teacher Qualifications

Correlation	Average Teacher Step	Median Teacher Step
QRIS PQ2 with VS-ORO Center Teacher Step	**0.70*** **	**0.72*** **

N-value varies for each correlation based on how many were observed. N=119.

*** Correlation is significant at the p<0.001 level (2-tailed)

Table 13h. Correlations between QRIS and SI Measures of Center Teacher Qualifications

Correlation	Correlation Coefficient
QRIS PQ2 with SI Teacher Step 7	**0.61*** **
QRIS PQ2 with SI Teacher Step 8	**0.62*** **
QRIS PQ2 with SI Teacher Step 9	**0.65*** **
QRIS PQ2 with SI Teacher Some College	**0.53*** **
QRIS PQ2 with SI Teacher Degree	**0.61*** **
QRIS PQ2 with SI Teacher ECE Degree	**0.53*** **

N=112.

*** Correlation is significant at the p<0.001 level (2-tailed)

Table 13i. Correlations between VS-ORO and SI Measures of Center Teacher Qualifications

Correlation	Average Teacher Step	Median Teacher Step
VS-ORO with SI Teacher Step 7	**0.57*** **	**0.57*** **
VS-ORO with SI Teacher Step 8	**0.54*** **	**0.55*** **
VS-ORO with SI Teacher Step 9	**0.58*** **	**0.57*** **
VS-ORO with SI Teacher Some College	**0.47*** **	**0.46*** **
VS-ORO with SI Teacher Degree	**0.52*** **	**0.49*** **
VS-ORO with SI Teacher ECE Degree	**0.47*** **	**0.46*** **

N=113.

*** Correlation is significant at the p<0.001 level (2-tailed)

PQ3: Center Aide and Certified Family Assistant Qualifications

Table 13j. Correlations between QRIS and VS-ORO Measures of Center Aide/CF Assistant Qualifications

Correlation	Center Aide		CF Assistant	
	Average Step	Median Step	Average Step	Median Step
QRIS PQ3 with VS-ORO Aide/Assistant Steps	**0.45****	**0.45****	**.56*****	**.54*****

N-value varies for each correlation based on how many were observed. For Centers N=43. For Certified Family N=43.

** Correlation is significant at the p<0.01 level (2-tailed)

*** Correlation is significant at the p<0.001 level (2-tailed)

Table 13k. Correlations between QRIS and SI Measures of Center Aide/CF Assistant Qualifications

Correlation	Center Aide	CF Assistant
QRIS PQ3 with SI Aide/Assistant Step 3	0.26^{+}	0.24
QRIS PQ3 with SI Aide/Assistant Step 5	**0.38***	**0.32***
QRIS PQ3 with SI Aide/Assistant Step 7	**0.35***	**0.33****
QRIS PQ3 with SI Aide/Assistant Some College or Degree	**0.33***	0.24
QRIS PQ3 with SI Aide/Assistant Degree	0.16	0.04
QRIS PQ3 with SI Aide/Assistant ECE Degree	0.23	0.07

N-value varies for each correlation based on how many were observed. For Centers Minimum N=38 and Maximum N=44. For Certified Family N=45.

$^{+}$ Correlation is nearing significance at the p<0.10 level (2-tailed)

* Correlation is significant at the p<0.05 level (2-tailed)

** Correlation is significant at the p<0.01 level (2-tailed)

Table 13l. Correlations between VS-ORO and SI Measures of Center Aide/CF Assistant Qualifications

Correlation	Center Aide		CF Assistant	
	Average Step	Median Step	Average Step	Median Step
VS-ORO Aide/Assistant Steps with SI Step 3	**0.59*** **	**0.59*** **	**0.55*** **	**0.55*** **
VS-ORO Aide/Assistant Steps with SI Step 5	**0.51** **	**0.51** **	**0.65*** **	**0.65*** **
VS-ORO Aide/Assistant Steps with SI Step 7	**0.46** **	**0.48** **	**0.67*** **	**0.68*** **
VS-ORO Aide/Assistant Steps with SI Some College/Degree	0.28	0.29	**0.42** **	**0.42** **
VS-ORO Aide/Assistant Steps with SI Degree	0.28	0.29	**0.43** **	**0.43** **
VS-ORO Aide/Assistant Steps with SI ECE-Related Degree	0.24	0.28	**0.47** **	**0.48** **

N-value varies for each correlation based on how many were observed. For Centers Minimum N=33 and Maximum N=35. For Certified Family N=44.

** Correlation is significant at the p<0.01 level (2-tailed)

*** Correlation is significant at the p<0.001 level (2-tailed)

PQ4: Correlations among Measures of Training

Table 13m. Correlations between QRIS and VS-ORO Measures of Training

Correlation	Center Staff	CF Staff	RF Staff
QRIS PQ4 with VS-ORO 18 hours or more	0.06	0.19^{+}	**0.30***
QRIS PQ4 with VS-ORO 20 hours or more	0.07	0.21^{+}	**0.33***
QRIS PQ4 with VS-ORO 24 hours or more	0.14	0.21^{+}	0.28^{+}

N-value varies for each correlation based on how many were observed. For Centers: Minimum N=120. For CF: N=81. For RF: N=45.

$^{+}$ Correlation is nearing significance at the p<0.10 level (2-tailed)

* Correlation is significant at the p<0.05 level (2-tailed)

Table 13n. Correlations between QRIS and SI Measures of Training

Correlation	Center Staff	CF Staff	RF Staff
QRIS PQ4 with SI Staff 18 hours or more	**0.43*** **	0.04	**-0.30***
QRIS PQ4 with SI Staff 20 hours or more	**0.51*** **	0.06	**-0.30***
QRIS PQ4 with SI Staff 24 hours or more	**0.58*** **	0**.26***	**-0.30***
QRIS PQ4 with SI Director/Provider 18 hours or more	0.10	0.08	**-0.30***
QRIS PQ4 with SI Director/Provider 20 hours or more	0.06	0.12	**-0.30***
QRIS PQ4 with SI Director/Provider 24 hours or more	0.08	**0.27***	**-0.30***
QRIS PQ4 with SI Teacher 18 hours or more	**0.41*** **	--	--
QRIS PQ4 with SI Teacher 20 hours or more	**0.49*** **	--	--
QRIS PQ4 with SI Teacher 24 hours or more	**0.55*** **	--	--
QRIS PQ4 with SI Aide/Assistant 18 hours or more	0.17	0.09	---
QRIS PQ4 with SI Aide/Assistant 20 hours or more	0.26^{+}	0.05	--
QRIS PQ4 with SI Aide/Assistant 24 hours or more	**0.32***	0.17	--

N-value varies for each correlation based on how many were observed. For Centers: Minimum N=57 and Maximum N=140. For CF: Minimum N=50 and Maximum N=68. For RF: Minimum N=44.

$^{+}$ Correlation is nearing significance at the p<0.10 level (2-tailed)

* Correlation is significant at the p<0.05 level (2-tailed)

*** Correlation is significant at the p<0.001 level (2-tailed)

Table 13o. Correlations between VS-ORO and SI Measures of Training

Correlation	Center Staff	CF Staff	RF Staff
VS-ORO 18 hours or more with SI Staff 18 hours or more	**0.39*** **	0.11	0.09
VS-ORO 20 hours or more with SI Staff 20 hours or more	**0.43*** **	0.19	0.06
VS-ORO 24 hours or more with SI Staff 24 hours or more	**0.47*** **	**0.37****	0.26^{+}
VS-ORO 18 hours or more with SI Director/Prov 18 hours or more	0.17^{+}	-0.12	0.09
VS-ORO 20 hours or more with SI Director/Prov 20 hours or more	**0.25****	-0.04	0.06
VS-ORO 24 hours or more with SI Director/Prov 24 hours or more	**0.33*** **	0.21^{+}	0.26^{+}
VS-ORO 18 hours or more with SI Teacher 18 hours or more	**0.21***	--	--
VS-ORO 20 hours or more with SI Teacher 20 hours or more	**0.25****	--	--
VS-ORO 24 hours or more with SI Teacher 24 hours or more	**0.34*** **	--	--
VS-ORO 18 with SI Aide/Assistant 18 hours or more	0.26^{+}	**0.33***	--
VS-ORO 20 with SI Aide/Assistant 20 hours or more	**0.40****	**0.37***	--
VS-ORO 24 hours or more with SI Aide/Assistant 24 hours or more	**0.45*** **	**0.33***	--

N-value varies for each correlation based on how many were observed. For Centers: Minimum N=57 and Maximum N=113. For CF: Minimum N=50 and Maximum N=68. For RF: Minimum N=35 and Maximum N=58.

$^{+}$ Correlation is nearing significance at the p<0.10 level (2-tailed)

* Correlation is significant at the p<0.05 level (2-tailed)

** Correlation is significant at the p<0.01 level (2-tailed)

*** Correlation is significant at the p<0.001 level (2-tailed

Measures of Administration and Business Practices.

The SI dataset also included measures of wages, benefits, teacher and provider retention, and accreditation. Data for these measures were collected directly from facilities at the time of their 2014 license renewal. In the Administration and Business domain, the QRIS dataset included a rating for benefits but not for retention, wages, or accreditation. The QRIS benefits rating (AB6) and SI benefits measure were not correlated, probably due to differences in the way the ratings and measures were calculated. SI included a count whereas QRIS granted a rating of 3 for providing 1 of 6 benefits, a 4 for 2 of 6, and a 5 for 3 of 6.

Correlations of QRIS AB Ratings with Additional SI Measures, Observed Quality, and Final Star Rating

Table 13p. Correlations among QRIS Administration and Business Practice Ratings and SI Measures for Centers

Measure	AB1	AB2	AB3	AB4	AB5	AB6	SI Benefits	SI Highest Wage	SI Lowest wage	SI Teacher Retention
AB1-Business records	**1**									
AB2—Working climate	**0.81*** **	**1**								
AB3-Staff evaluations	**0.69*** **	**0.71*** **	**1**							
AB4-Professionalism	**0..82*** **	**0.86*** **	**0.68*** **	**1**						
AB5-Program evaluation	**0.75*** **	**0.71*** **	**0.62*** **	**0.69*** **	**1**					
AB6-Benefits	**0.78*** **	**0.75*** **	**0.66*** **	**0.82*** **	**0.65*** **	**1**				
SI Benefits	-0.11	0.05	-0.16	-0.15	-0.14	-0.18	**1**			
SI Highest Wage	**0.31****	**0.42*** **	**0.44*** **	**0.31****	**0.33****	**0.35****	0.21^{+}	1		
SI Lowest Wage	**0.28***	**0.32****	**0.34****	**0.30***	**0.33****	**0.36****	-0.00	0.63*** **	**1**	
SI T Retention	0.16	0.12	**0.23***	**0.19***	**0.22***	0.11	-0.15	0.20^{+}	**0.30****	
Accreditation	**0.25****	**0.27****	**0.35*** **	**0.25****	**0.29****	**0.24***	-0.03	0.13	0.21^{+}	0.12

N-value varies for each correlation based on how many were observed. Minimum N=69 and Maximum N=119.

$^{+}$ Correlation is nearing significance at the p<0.10 level (2-tailed)

* Correlation is significant at the p<0.05 level (2-tailed)

** Correlation is significant at the p<0.01 level (2-tailed)

*** Correlation is significant at the p<0.001 level (2-tailed

Table 18a. Correlations among Personnel Measures and Observed Quality in Centers.

Construct	Measure	CLASS Domains		
		Emotional Support	Instructional Support	Organized Classrooms
Wages	Teacher Lowest Wage	0.15	**0.32****	0.20+
Retention	% Teachers Retained	0.10	0.09	0.09
Benefits	Benefits	0.10	0.16	0.16
Accreditation	Accreditation	-0.02	0.12	-0.06

N-value varies for each correlation based on how many were observed. Minimum N=75 and Maximum N=141.
+ Correlation is nearing significance at the p<0.10 level (2-tailed)
** Correlation is significant at the p<0.01 level (2-tailed)

Table 18b. Correlations among Personnel Measures and Observed Quality in Certified Family.

Source	Measure	Class Domain		
		Emotional Support	Instructional Support	Organized Classrooms
Retention	Provider Retention	0.02	0.17	0.18
Accreditation	Accreditation	0.06	0.02	0.09

N-value varies for each correlation based on how many were observed. Minimum N=62 and Maximum N=77.

Table 18c. Correlations among Personnel Measures and Observed Quality in Registered Family.

Construct	Measure	CLASS DOMAIN		
		Emotional Support	Instructional Support	Organized Classrooms
Retention	Provider Retention	-0.09	0.09	**0.37***
Accreditation	Accreditation	0.02	-0.02	0.03

N-value varies for each correlation based on how many were observed. Minimum N=45 and Maximum N=59.
* Correlation is significant at the p<0.05 level (2-tailed)

Table 19a. Correlations among Personnel Measures and Final Star Rating in Centers.

Construct	Measure	
Wages	Teacher Lowest Wage	**0.38*** **
Retention	% Teachers Retained	0.16
Benefits	Benefits	0.05
Accreditation	Accreditation	**0.41*** **

N-value varies for each correlation based on how many were observed. Minimum N=82 and Maximum N=141.
*** Correlation is significant at the p<0.001 level (2-tailed

Table 19b. Correlations among Personnel Measures and Final Star Rating in Certified Family.

Source	Measure	
Retention	Provider Retention	0.09
Accreditation	Accreditation	**0.34** **

N=77.
** Correlation is significant at the p<0.01 level (2-tailed)

Table 19c. Correlations among Personnel Measures and Final Star Rating in Registered Family.

Source	Measure	
Retention	Provider Retention	-0.03
Accreditation	Accreditation	0.08

N=59