**Qualistar Colorado: Supporting Research on Early Care and Education Quality**

*Data on early care and education quality, drawn from the Qualistar Colorado Quality Rating and Improvement System (QRIS), and child outcome data drawn from the RAND Corporation’s evaluation of Colorado’s QRIS have been published in at least 10 highly respected journals and have been used to support two dissertations. Data have also been used in three successful federal grant competitions; one sponsored by the National Institute of Health (NIH) and two sponsored by the federal Office of Planning, Research, and Evaluation (OPRE), Administration for Children and Families, and an additional proposal using Qualistar data has been submitted to the Institute of Educational Sciences (IES). In turn, results of these studies have been used to: (1) inform how to measure quality in the former Qualistar Rating in a valid and reliable manner and to provide technical assistance to other states in the development of their rating approaches, (2) to inform statistical methods for QRIS validation studies, (3) to guide an intervention to improve the caregiving consistency and teacher-child attachment relationships in an infant-toddler child care intervention, and (4) to inform measurement approaches in a family child care professional development study.*

*Annotated Bibliography*

 *Research Supported by Qualistar data*

**Measuring Quality**

The following studies have used data collected by Qualistar or from the RAND Corporation’s evaluation of the Qualistar Rating to address how to reliably and cost effectively measure quality in high stakes contexts such as in QRISs. Reliability is of particular concern because if dimensions of quality are not measured well and consistently across programs, then their connections to child outcomes will be limited and policymakers will not have accurate information about important levers that may improve outcomes of interest.

Examining the Number of Items Needed to Observe in the family of Environmental Rating Scales

This group of studies examined whether there are a reduced number of items that can be sampled on the Early Childhood Environment Rating Scale-Revised (ECERS-R), the Infant Toddler Environment Rating Scale-Revised (ITERS-R), and the Family Child Care Environment Rating Scale-Revised (FCCERS-R) that can be used to obtain an overall score similar to what would be earned if the entire scales were administered so as to achieve cost savings in ratings.

***Examining the Psychometric Properties of the Early Childhood Environmental Rating Scale-Revised*,** by Michal Perlman (University of Toronto, formerly of RAND), Gail Zellman (RAND), and Vi-Nhuan Le (NORC at the University of Chicago, formerly of RAND)

This study found that there was only one overarching dimension of quality measured in the ECERS-R, as opposed to the seven dimensions of quality that are purported by the authors. Therefore, the authors selected 24 items which were easy and cost effective to administer and 10 items on the ECERS-R that were deemed by ECE experts to be most important to support child learning and development. The authors found that they could use both types of item sampling strategies to achieve an overall ECERS-R score similar to as if they had administered the entire scale. They then found that higher overall ECERS-R scores achieved by the two sampling techniques were related to lower classroom ratios, and to higher teacher education levels; suggesting that a smaller subset of items on the ECERS-R can be used to achieve measurement efficiency in the learning environment domain of a QRIS.

***Examining the Psychometric Properties of the Infant and Toddler Environment Rating Scale-Revised in High Stakes Contexts*,** by Rosanne Biscaglia (University of Toronto), Michal Perlman, Diana Schaack (San Diego State University, formerly of Qualistar), and Jennifer Jenkins (University of Toronto)

This study examined whether a subset of items on the ITERS-R could be used to more efficiently measure infant and toddler child care quality. The authors created a randomly selected subset of 10 items, a set of 10 items considered by experts to be important to child learning and development, and a set of 10 easy to administer items. They found that the randomly selected and expert endorsed items were a reliable representation of the overall scale and that these two sampling techniques demonstrated slightly better associations with teacher education levels and classroom ratios than the entire ITERS-R scale. The best assessment of quality was demonstrated by the shortened subset that included items that assess both structural and process features of quality.

***Examining the Factor Structure of the Family Child Care Environment Rating Scale-Revised,*** by Diana Schaack, Vi-Nhuan Le, and Claude Messan Setodji (RAND)

This study found that there were three underlying dimensions of quality measured in the FCCERS-R; an Activities/Materials scale, a Language/Interaction scale, and an Organization of the Day/Physical Environment scale. The authors looked at a number of ways in which to sample a smaller group of items on the FCCERS-R to determine if a smaller subset of items would achieve a similar overall score as administering the entire scale. They found that sampling 2-3 items in each of the three scales: Activities/Materials, Language/Interactions, and Organization provided a cost effective way to reduce the FCCERS-R in a way that would yield similar results to the entire scale. They also determined that this method of sampling items demonstrated positive relationships to family child care provider education levels.

How Many Classrooms to Sample when Administering the ERS

This research study, “Capturing the Heterogeneity in Quality within Early Care and Education Programs Serving Preschool Children (2009-2012) was funded by a grant from the federal Office of Planning, Research, and Evaluation to Gail Zellman, Lynn Karoly, and Michal Perlman. Broadly, this grant sought to identify: (1) how to combine measures of individual teachers to best capture quality at the classroom level, (2) whether quality should be measured at the staff, classroom or center level, and (3) whether there are ways to achieve measurement efficiency. To date, one peer-reviewed article has been published from this study.

***Understanding Variation in Classroom Quality Within Early Childhood Centers: Evidence from Colorado's Quality Rating and Improvement System,*** by Lynn Karoly (RAND), Gail Zellman, and Michal Perlman

This sought to identify the number of classrooms that need to be observed (e.g., how many need to be sampled) using the Environment Rating Scales to obtain an accurate account of the overall quality of all classrooms. The study first found that there were significant differences in classroom quality within centers, therefore they experimented with different ways in which to sample classrooms. They found that for centers ranging from four rooms in total to eight rooms in total, the best sampling approach was to select at least two classrooms of the same age group to observe. They also found that for centers with less than eight classrooms, observing at least 50% of the classrooms in the same age range produced the most accurate approximation of overall quality.

How Best to Measure Classroom Ratios

***Measuring Child-Staff Ratios in Child Care Centers: Balancing Effort and Representatives,*** by Vi Nhuan-Le, Gail Zellman, Michal Perlman, and Laura Hamilton (RAND)

This study found that there was a great deal of variation in classroom group sizes and ratios within a center and even within the same age group. For example, one classroom could be in compliance with licensing standards, while another classroom may not be in compliance. Consequently, they concluded that averaging ratios within a QRIS rating component was not a recommended strategy, since it could be very misleading; especially to parents whose children are in the out of compliance classrooms. Instead they endorsed a scoring approach were a center is awarded points based on the ratios in the classroom with the highest ratios. They also found that measuring ratios just in the morning during Environment Rating Scale observations underestimates the actual ratios children experience and that ratios need to be collected periodically throughout the entire day to obtain an accurate account of what children experience. This study also points out that administrative data, for example licensing compliance records, may not be an accurate account of the ratios children experience in their classrooms.

Conceptualizing Parent Involvement Criteria

***Reconceptualizing the role of parents' involvement in their children's child care providers*,** by Michal Perlman and Gail Zellman

This paper discusses the conceptualization and measurement of parent involvement and family partnerships in quality rating and improvement systems and discusses the difficulty of measuring this construct. It presents data on several parent involvement measures, including one used by the National Association for the Education of Young Children, the Parent Caregiver Relationship Scale (Elicker et al., 1997) items from the National Longitudinal Survey of Youth, the Early Childhood Environment Rating Scale—Revised (ECERS-R) Parent items, and the Qualistar Family Partnership survey. Across these measures parents uniformly describe child care centers as welcoming and supportive even when other quality measures reveal significant problems with the center. Directors display similar positivity when reporting on their own parent involvement efforts. The Qualistar Family Partnerships measures produce similar positive ratings as other parent involvement measures; however the Qualistar measure was more strongly associated with other assessments of child care quality, including the ECERS-R, the ITERS-R, classroom ratios, and teacher education than the other parent involvement measures indicating it may be important to include this component in QRISs.

***Where to Set Rating Levels on QRIS indicators***

This project, *Identifying Thresholds in Child Care Quality Measures*, was funded by the National Institute of Health and was led by Claude Messan Setodji and Vi-Nhuan Le. The study sought to explore new statistical methods using national datasets to identify cut-points along the quality continuum related to better child learning and development.

***Identifying baseline and ceiling thresholds in the Qualistar Quality Rating and Improvement System,*** by Vi-Nhuan Le, Diana Schaack, and Claude Messan Setodji.

This study reanalyzed Qualistar child outcome data using a novel statistical technique; one that does not assume that as quality gets higher, child outcomes get continuously better. It instead looks at whether there are certain levels of quality that must be reached before better child outcomes can occur. The study found that once quality reached a 3.40 on the ECERS-R, better child cognitive outcomes were observed, and then again once ECERS-R quality reached a level of 4.60.  With a bigger sample, it appears that another cut-point related to better cognitive development may be at 6.00 on the ECERS-R. The study also found that better ECERS-R quality was related to better social outcomes and that quality did not have to reach a particular level before better social outcomes were observed. Together, this provides evidence for the validity of Qualistar's star levels on the learning environment component. The study also found evidence for the validity of Qualistar's first threshold on the teacher education component --but once teacher's had completed more than 12 credits in ECE, there were no relationships to better ECERS-R scores.  The study also indicates that until teachers have at least 15 years of experience, no relationships to better classroom quality can be observed; suggesting that Qualistar may want to raise the ceiling on the experience level for this component of the rating system. Finally, the highest ratio point level on the Qualistar Rating (a ratio of 1:8 children) was related to better ECERS-R scores; but lower point levels of the Qualistar rating ratio component were actually related to worse outcomes; suggesting that Qualistar may want to readjust their ratio levels.

Assessing the Continuity of Children’s Experiences in Child Care Settings

This set of studies was funded by a grant from the National Institute of Health to Vi-Nhuan Le and Claude Messan Setodji and it investigated the amount of daily transitions that children experience in their child care settings and its effects on their social, emotional, and cognitive development using data from the evaluation of the Qualistar QRIS.

***Examining the associations between children’s daily caregiving discontinuity experiences and children’s social-emotional outcomes,*** by Vi-Nhuan Le, Diana Schaack, and Claude Messan Setodji

This study examined the prevalence and patterns of daily teacher and child movement between classrooms (also called accordion grouping), the characteristics of teachers and children who were more likely to move between classrooms on a daily basis, and the associations between children’s and teachers’ rate of daily movement between classrooms with children’s social-emotional outcomes. A moderate to high prevalence of child and teacher movement between classrooms was observed, namely 29% of teachers moved to different classrooms for an average of two hours each day, and 83% of children where moved out of their assigned classroom for an average of an hour outside of their classrooms each day. Children who were younger, considered solitary, and who had been enrolled in their classroom for shorter periods of time were less likely to transition between classrooms. Children who moved more each day had more conflictual relationships with teachers and were less likely to like going to their centers. However, children who moved classrooms more tended like their peers more and their peers indicated liking them more as well. Nonetheless, concerted efforts may need to be made in policy to reduce the amount of discontinuity in children’s caregiving and learning experiences.

***Accounting for movement between childcare classrooms: Does it change teacher interpretation effects,*** by Claude Messan Setodi, Vi Nhuan Le, and Diana Schaack

Currently a debate exists in the field about the value of an early childhood teacher BA degree. Some research has found support that a BA is related to better teaching and child learning, while others have not. This study first found that there was a large prevalence of children and teachers moving in and out of different classrooms each day and that this movement has important implications for the relationship between teacher education and child outcomes. That is, the authors found that it was important to take into account the actual amount of time a child spent with a teacher with a BA, and once that was accounted for, children who spent more time with teachers with a BA had better school readiness skills than children who spent less or no time with a teacher with a BA. This study also noted that many children were assigned to teachers with a BA, but because of the frequent reorganization of children and teachers throughout the day, children often did not spent much time with their assigned teacher. This study suggests that quality improvement and teacher education interventions must be aimed at all teachers within a center, since children tend to spend substantial time with teachers other than the ones they are assigned to.

QRIS as a Quality Improvement Intervention

This study was funded by a grant from the federal Office of Planning, Research, and Evaluation.

**Substantive or symbolic stars: Quality rating and improvement systems from a new institutional lens,** by Kate Tarrant and Luis Huerta

In this study, the authors examined QRIS participants’ perceptions of the Qualistar Rating and the changes that they made in their practices as a result of the rating. Programs at one end of the spectrum were classified as those that felt positively about the Qualistar QRIS. With few exceptions, teachers in these groups noted that their participation had a substantive impact on their practice. These centers had low initial ratings, and were not affiliated with any other quality initiative or program; therefore Qualistar introduced them to higher standards and provided them a roadmap to get to higher quality. In the middle of the spectrum were those who were ambivalent about the QRIS. These teachers worked in centers with high initial ratings and these teachers noted that the rating raised their awareness to programmatic areas that they may have overlooked (e.g., science), but were irritated by the rigidity of the ECERS-R, but felt that overall the rating was aligned with their center’s educational approach. Therefore they made some modest changes and felt appreciative of the financial support that allowed them to make the changes. They also noted how the changes made might not be sustained because of other competing and often at odds policy demands (e.g., Head Start standards). At the far end of the spectrum were teachers in centers who held negative views of the ratings, some of whom outright rejected the rating. These were teachers in centers with high quality ratings who felt that the rating was burdensome and unhelpful, and if they made changes, they were symbolic and only for the day of the rating. This study provides some insight on how to target ratings and that more sophisticated instructionally focused measures may be needed as Colorado’s programs improve along the quality spectrum.

Dissertations using Qualistar Data

**Substantive or symbolic stars: A mixed methods analysis of quality rating and improvement systems through a new institutional lens**, by Kate Tarrant (chair Lynn Kagan)

This study suggests that programs participating in Colorado’s QRIS experienced inconsistent improvements to process quality. The quantitative results indicated that programs most likely to improve in process quality where low initial scoring programs. The qualitative results suggested that centers with highly qualified teachers that were affiliated with policy initiatives (i.e., Head Start or pre-kindergarten) or had high ratings reported modest or no process quality improvements. The key factors that promoted process quality improvements were: coaching, professional development, program administrator leadership, and curriculum implementation. Several of these factors were not directly part of the QRIS improvement strategy. With respect to structural quality improvement, the study found that programs made more substantial improvements to structural quality. The quantitative data showed that structural quality improvements were associated with low-quality scores. Within the qualitative sample, respondents reported they made structural quality improvements at all sites. The key factors affecting these changes were coaching, funding, and the high stakes associated with the rating.

**A descriptive-comparative study of professional development and observed quality in early care and education programs**, by Rebecca Romeyn (chair Toni Linder)

To contribute to the advancement of accurately describing and understanding effective early childhood teachers' professional development experiences in terms of content, quantity, intensity, and duration, a descriptive-comparative design was used to explore how the professional development experiences of teachers in 10 good quality classrooms (as measured by the Early Childhood Environment Rating Scale-Revised [ECERS-R]) differed from the professional development experiences of teachers in 10 poor quality classrooms serving children ages 2 1/2 to 5 years old. Teachers in the good quality cohort were more likely to have completed course content specific to early care and education, including specific combinations of coursework in conjunction with supervised practical experience. The intensity and duration of teachers' exposure to course content was also greater in this cohort. Furthermore, in describing differences, new categories and definitions emerged, including piecemeal and comprehensive education, piecemeal and comprehensive training, content clustering, solo teaching, and hierarchical structuring of staff.