



Contents lists available at ScienceDirect

## Early Childhood Research Quarterly



# Quality Rating and Improvement Systems: Validation of a local implementation in LA County and children's school-readiness

Sandra L. Soliday Hong<sup>a,\*</sup>, Carollee Howes<sup>b</sup>, Jennifer Marcella<sup>b</sup>, Eleanor Zucker<sup>b</sup>, Yiching Huang<sup>b</sup>

<sup>a</sup> University of North Carolina, Chapel Hill, United States

<sup>b</sup> University of California, Los Angeles, United States

### ARTICLE INFO

#### Article history:

Available online xxx

#### Keywords:

Quality Rating and Improvement Systems  
Child care quality  
School-readiness

### ABSTRACT

As states are actively participating in the Race to the Top Early Learning Challenge Grant, research that validates the translation of child care quality measures for policy purposes is required. This paper presents results from a two-part study on QRIS ratings: (1) Study 1 examines the concurrent validity of a QRIS, and (2) Study 2 simulates the QRIS scores with a secondary dataset to predict child outcomes. Study 1 – the QRIS descriptive study – presents the pilot-study data from the L.A. County QRIS collected between 2009 and 2012. Data from 254 early childhood programs (98 family child care homes and 156 centers with 331 classrooms) indicate that individual quality rating indicators do contribute to an overall score, and the simulated scores were related to external measures of child care quality. Study 2 – the QRIS simulation study – included 223 low-income 3- and 4-year-old children in 101 early care and education classrooms/programs; one-quarter were dual language learners. Continuous measures of child care quality were positively associated with children's school-readiness. However, after quality measures were combined and scored into QRIS ratings, ratings were no longer associated with increases in cognitive/academic and social child outcomes across the school year. The lack of an association between QRIS ratings and child outcomes is consistent with findings from other studies and warrants further examination of QRIS ratings and their effect on children's learning.

© 2014 Elsevier Inc. All rights reserved.

### Introduction

Previous research illustrates the well-established link between the quality of early care and education and children's school-readiness (Howes et al., 2008; NICHD Early Child Care Research Network, 2002, 2003; Shonkoff & Phillips, 2000). For example, recent meta-analysis demonstrated that higher quality early care and education relates to higher language, academic, and social skills, and fewer behavior problems (Burchinal, Kainz, & Yaping, 2011). Taken together, these studies show that high-quality early care and education in the form of intensive teacher-child interactions relates to children's school readiness outcomes.

Due to the significance of quality in early care and education programs, states and counties have established Quality Rating and

Improvement Systems (QRIS) over the last decade, and recently with support from the Federal Race to the Top Early Learning Challenge Grant (RTT-ELC). QRIS collect information about the quality of ECE programs, convert scores from continuous quality measures into categorical quality indicator ratings, and combine the categorical quality ratings into overall program-level ratings which provide benchmarks that give programs information relevant to improving their quality (Tout et al., 2010; Zellman & Fiene, 2012). These program ratings are made publicly available to parents and policymakers, with the intention of incentivizing voluntary improvement (Schaack, Tarrant, Boller, & Tout, 2012; Zellman & Perlman, 2008). QRIS are typically developed through a constituent-based process intended to improve quality through consensus on outcomes considered to be important by the local QRIS decision-making committees (Zellman & Fiene, 2012). As such, QRIS may have broad and diverse goals such as improving school readiness outcomes for children, professionalizing the early care and education workforce, enhancing family outcomes, or improving parent's knowledge of local programs. Unfortunately, relatively little attention has been paid to examining the empirical basis for specific implementations of QRIS and the link between QRIS indicators considered to be

\* Corresponding author at: FPG Child Development Institute, University of North Carolina, Chapel Hill, 521 S. Greensboro St., Chapel Hill, NC 27599-8185, United States. Tel.: +1 919 966 9513.

E-mail address: [sandy.hong@unc.edu](mailto:sandy.hong@unc.edu) (S.L. Soliday Hong).

important by the stakeholders with the intended outcomes. The current paper describes and measures the potential impact of a local QRIS in Los Angeles County (LA QRIS) to inform future states' policy making with a particular focus on children's outcomes.

#### *Common QRIS indicators*

Due to the diverse stakeholders who contribute to the development of QRIS, a wide array of quality indicators may be included in QRIS ratings. Many QRIS are structured so that basic health and safety standards related to child care licensing form the anchor at the lowest end of the rating structure, and accreditation by a nationally recognized early childhood agency guide the requirements for the top rating, although more variability exists at the bottom end of the scales than at the higher levels (Caronongan, Kirby, Malone, & Boller, 2011). Ratings vary in the criteria that they include, but some of the most common elements are: licensing compliance, learning environments, staff qualifications, efforts to strengthen and form partnerships with families, administration and management, and accreditation (Tout et al., 2010). Additional criteria include the use of research-based curricula, teacher–child ratios and group size, child assessment, health and safety, cultural and linguistic diversity, provisions for children with special needs, and community involvement.

These quality indicators are typically measured by observational measures of children's experiences with their teachers and peers which are scaled with continuous measures, and checklist measures of structural quality and program engagement with families and the community, both of which need to be translated into points or categorical quality ratings depending on the structure of the individual QRIS. The underlying assumption of QRIS ratings is that higher ratings will have a stronger association with desired outcomes of the system, like children's school-readiness. To date, little research is available to guide states in making decisions regarding cut-points in these quality measures and the concurrent validity of ratings that combine differing thresholds of varying measures, therefore, sharing descriptive information among states on the cut-points and the relation of those cut-points to QRIS related goals is valuable in guiding state QRIS decision-making.

While there is growing consensus within the research community that teacher–child interactions are the main predictor of school-readiness outcomes, QRIS generally contain many more elements than teacher–child interactions which may relate to child outcomes or other goals related to ECE systems building (Caronongan et al., 2011; Elicker, Langill, Ruprecht, & Kwon, 2007; Schaack et al., 2012; Tout et al., 2010). For example, an older research literature linked higher teacher education levels to teacher practice which was assumed would translate into better child outcomes (Howes, 1997; Phillips, Mekos, Scarr, McCartney, & Abbott-Shim, 2000; Phillipsen, Burchinal, Howes, & Cryer, 1997). As a result, many QRIS include teacher preparation through both formal education and certification. To further complicate the development of QRIS ratings, early care and education regulations in the United States are differential depending on the age of the child and the auspice of the program, therefore group sizes, ratios, and even environmental rating systems such as the Environmental Rating Scales (ERS; Harms, Cryer, & Clifford, 2007; Harms, Clifford, & Cryer, 1998; Harms, Cryer, & Clifford, 2000) are differential by age group as well as by form of care setting (family child care and center based care), creating a complex set of guidelines for calculating QRIS ratings.

#### *QRIS ratings and school-readiness*

Some of the strongest predictors of children's gains in pre-academic skills are teacher instructional interactions, warmth and

responsivity, so we would expect that quality ratings based on measures of teacher–child interactions would be of particular relevance in improving children's outcomes within QRIS systems (Howes et al., 2008). Teachers' instructional interactions with children predict children's academic and language outcomes, while teachers' emotional interactions with children predict children's social skills (Mashburn et al., 2008). Additionally, teacher classroom practices (i.e., intensity of involvement, child centeredness, and developmental appropriateness) relate to children's language and academic skills, while teacher–child closeness relates to children's cognitive and social skills (Peisner-Feinberg et al., 2001). Although a predictive relation has been established between these continuous measures of teacher–child interactions and children's development and school-readiness, little research has been done to translate these continuous predictors into categorical ratings or points as are frequently utilized in QRIS ratings and to examine if these relations persist. More information is needed about various QRIS scoring configurations and children's school-readiness.

A recent study explicitly examined variations in scoring approaches used by state QRIS by simulating the scoring configurations of nine states using existing data from a large study of public pre-k (Sabol, Soliday Hong, Pianta, & Burchinal, 2013). This study found that overall QRIS scores were generally not predictive of gains in children's school-readiness skills across the school year in highly rated programs regardless of the type of scoring approach adopted by states with complex quality indicators. However, examination of variation of ratings across state criteria showed that states that streamlined their ratings to focus on teacher–child relationships were more predictive of child outcomes than rating systems that were inclusive of additional and at times loosely related criteria. This secondary study reflects the results of state QRIS evaluations that report mixed results for the predictability of QRIS ratings on child outcomes (Barnard, Etheridge Smith, Fiene, & Swanson, 2006; Child Trends, 2010; Thornburg, Mayfield, Hawks, & Fuger, 2009).

#### *Implementing a local QRIS*

The LA County QRIS (LA QRIS) incorporates many of the common quality indicators that comprise QRIS ratings across the country. For this reason, and due to the diverse nature of the child and family population in combination with historically low levels of quality observed in ECE programs, the LA QRIS was examined as an example of QRIS in the United States. Additionally, although the LA QRIS is a county-level project, it is comparable to other state-level QRIS due to the large number of licensed child care providers, which is higher than the number of providers in 46 states (Los Angeles County Policy Roundtable for Child Care, 2006).

A potentially unique feature of the LA QRIS is that it grew out of several decades of collaboration between a local university partner and a group of stakeholders, including a five-year process of developing and pilot testing of the QRIS rating matrix, and the LA QRIS is inclusive of programs that serve children from birth through kindergarten entry in multiple ECE auspices including centers and FCCs that are housed, regulated, and funded by various agencies. However, in a similar process to other counties and states, a policy planning group comprised of a variety of stakeholders involved in the child care field were also involved in the consensus-focused process which incorporated alignment of QRIS ratings with state and local regulations and paperwork requirements: from practitioners, to child and family service providers, to child and family program developers, to local/county level policymakers, and university researchers.

The goals of the consensus-driven process were to create a rating system that would have face validity to the local community, that the ratings would be meaningful and useful to them for quality improvement, and that the ratings would align with other

regulatory agencies to reduce undue reporting burden on participating programs. The impetus from the research side was to create a rating system that was parsimonious and used reliable instruments that have been linked to child outcomes to measure quality. From the policy side, there was a strong motivation to include indicators which helped push practitioners in the child care field toward higher standards of professionalization with the goal of building early care and education systems building (e.g., staff qualifications based on the permit and degree system, working conditions; Schaack et al., 2012). Also of import for policymakers was to help programs develop areas they might not otherwise delve into, such as working with children with special needs, and improving outreach with the goal of strengthening families.

The LA QRIS has six indicators of quality that are similar to many other states as described above: (1) regulatory compliance, (2) teacher–child relationships, (3) learning environment, (4) inclusion of children with special needs, (5) staff qualifications and working conditions, and (6) family and community connections which are differential contingent on child age and auspice. Additionally, the LA QRIS is voluntary and inclusive of all licensed ECE programs that serve children from birth through kindergarten entry. As such, information about the design of the program for multiple age groups and auspices, and the relation between the LA QRIS indicator and overall ratings and child outcomes for children approaching kindergarten entry might be instructive for other states as they seek to gather information about the variety of options to select measures, design quality indicators, and examine various scoring methods.

Similar to many states, the resulting QRIS was intended to be evidence-based and to provide a link between certain indicators within the rating system to positive influences on children's development (Caronongan et al., 2011; Elicker et al., 2007). QRIS typically use one of three different strategies as a guideline for combining QRIS indicator ratings into overall QRIS scores: (1) block systems where programs must meet all of the criteria related to quality indicators before they can qualify for the next overall quality rating level, (2) point systems which assign points associated with benchmarks on the quality indicators, and (3) hybrid approaches which use some combination of the two approaches with varying criteria (Tout et al., 2010). The LA QRIS scoring uses a hybrid block system that gives larger importance (through use of weights) to items that have been more strongly linked to child outcomes in the research literature.

Prior to conducting a pilot study of the rating criteria from 2009 to 2012, QRIS ratings were simulated using secondary data from research projects conducted in the local area which were then utilized to simulate the relation between quality measures and child outcomes. The current study takes this one step further by simulating the ratings as they evolved throughout the QRIS pilot, and comparing children's learning in programs that received high ratings with programs that received lower ratings in an attempt to assess the impact of the combined ratings on children's development.

### Summary

This paper presents the results from two studies examining the QRIS implementation in Los Angeles County. In Study 1, we ask the following questions: (1) Does the LA QRIS scoring criteria have concurrent validity across measures and quality indicators? and (2) Do children of different ages, and in varying ECE programs receive ECE services that vary systematically in the voluntary QRIS? Accordingly, Study 1 is designed to examine the concurrent validity of this QRIS by testing its internal consistency using data collected by the system during the 3-year pilot phase, and to describe QRIS scoring by auspice and child age. The second study asks, (1) Does the

concurrent validity of the LA QRIS replicate with a secondary dataset collected in the same community? (2) Do the LA QRIS indicator and overall ratings have predictive validity for children's school-readiness? As such, the Study 2 simulation builds upon the Study 1 QRIS descriptive study by replicating and testing both the concurrent and predictive validity of the QRIS ratings with external quality measures using a secondary data analysis as child outcome data were not collected during the pilot phase of the QRIS. Using an existing dataset from Los Angeles County, we replicated the QRIS indicators and overall scores to see how well they predict child outcomes. These studies will inform gaps in the literature on the validity of QRIS scoring, and test the underlying assumption that higher QRIS ratings relate to children's school-readiness.

### Study 1: QRIS descriptive study

#### Study 1: method

#### Study 1: participants

There are just over 3000 (3070) licensed child development centers and nearly 11,000 (10,962) licensed family child care (FCC) homes in Los Angeles County. Combined, these programs have an estimated capacity to care for 226,000 children ranging in age from six weeks to five years (Los Angeles County Policy Roundtable for Child Care, 2006). LA QRIS was designed to be inclusive of all licensed programs serving children birth to five years of age, including FCCs, and private center based care, in addition to publicly funded programs such as Head Start and state pre-k programs run by school districts.

The LA QRIS is a voluntary program, and centers and family child care homes were recruited to participate by an LA County early childhood policy agency, with special recruitment occurring in nine areas of LA County that were determined to be of "high need" of quality early care and education services based on a lack of available licensed programs compared to census records of children residing in the area, or due to high levels of poverty and socioeconomic risk in those communities. Outreach was conducted through local early childhood agencies in those communities. Data on 254 early childhood programs who volunteered to participate in the QRIS were collected during the three-year pilot phase of the QRIS from 2009 to 2012 with 98 family child care homes, and 156 centers with 331 classrooms. Data were collected using research-quality standards of reliability, independent observations, data collection, and data management in partnership with a local university, and QRIS scores were calculated and issued by the university. This QRIS data set includes varied child care auspices and age groups from birth through pre-kindergarten.

Programs were rated on a scale of one-to-five on six domains: compliance with basic licensing standards (pass/fail), teacher–child relationships, learning environments, identification and inclusion of children with special needs, staff qualifications and working conditions, and family and community partnerships. Infants are defined as children from birth to 18 months, toddlers are between 18 and 30 months, and for the purposes of this paper preschool-age is defined as children 30 through 48 months (or two years before kindergarten), and pre-k includes children from 48 months through kindergarten entry (or the year before kindergarten). In an attempt to streamline the system, and use parallel assessments with various age groups and auspices, the Early Childhood Environment Rating Scale (ECERS-R; Harms et al., 1998) factor scores (Cassidy, Hestenes, Hedge, Hestenes, & Mims, 2005) were adapted for use with the Infant/Toddler Environment Rating Scale (ITERS-R; Harms et al., 2000), and the Family Child Care Environment Rating Scale (FCCERS-R; Harms, Cryer, & Clifford, 2007), and the Adult Involvement Scale (AIS; Howes & Stewart, 1987) was utilized across

**Table 1**  
Descriptive statistics of observational measures, and correlations among observed classroom characteristics, quality, and QRIS scores.

	Median child age (Mos.)	Group size	Teacher–child ratio	Adult inv. scale	ERS		CLASS			
					Teach. & interact.	Prov. for learn	Overall score	Emot. climate	Instr. climate	Class org.
<b>Study 1: QRIS descriptive study</b>										
<i>n</i>	429	429	429	429	429	429	–	–	–	–
<i>M</i>	42	14.78	.27	4.42	5.94	4.45	–	–	–	–
<i>SD</i>	–	7.42	.19	.72	.82	.96	–	–	–	–
Teacher–child relationships	.17**	.23***	–.32***	.75***	.46***	.50***	–	–	–	–
Learning environment	.14*	.23***	–.22***	.47***	.79***	.82***	–	–	–	–
Special needs	.03	.20***	–.26***	.09	.28***	.33***	–	–	–	–
Staff qualifications & working conditions	.04	.22***	–.24***	.17**	.25***	.22***	–	–	–	–
Family & community	.16**	.31***	–.37***	.27***	.33***	.39***	–	–	–	–
Overall score	.18**	.35***	–.36***	.24***	.59***	.62***	–	–	–	–
<b>Study 2: QRIS simulation study</b>										
<i>n</i>	91	99	98	99	94	–	98	98	98	98
<i>M</i>	38.39	18.18	.16	5.25	5.21	–	4.84	6.06	2.17	4.56
<i>SD</i>	3.07	7.12	.05	.49	1.40	–	.39	.42	.62	.53
Teacher–child relationships	.06	–.09	–.12	.38***	.11	–	.28**	.24*	.21*	.21*
Learning environment	.35***	–.17	.02	.37***	.97***	–	.60***	.63***	.47***	.26*
Special needs	–.22*	–.07	–.11	–.25*	–.29*	–	–.23*	–.34**	–.07	–.07
Staff qualifications & working conditions	–.04	.07	–.05	.13	.13	–	.15	.05	.26*	.10
Overall score	.10	.15	.06	.01	.08	–	.06	.03	.15	–.02

\* *p* < .10.  
 \* *p* < .05.  
 \*\* *p* < .01.  
 \*\*\* *p* < .001.

these groups. More information about the use of these measures follows.

The scoring of the LA QRIS uses a hybrid block system which gives larger weights within indicators to items that have been more strongly linked to child outcomes in the research literature. For example, the staff qualifications and working conditions indicator scoring favors staff qualifications over the other indicators geared toward workforce development. In two other quality indicators—how well programs serve children with special needs, and connections with families and the community—the lack of research linking these areas to child outcomes necessitated that the weighting decisions be guided by consensus rather than evidence from the extant research. The QRIS criteria are intended to align with basic licensing standards at the entry level, with Title 5/California Department of Education standards at the mid-level, and with National Association for the Education of Young Children Accreditation (NAEYC) or National Association for Family Child Care (NAFCC) standards at the top level.

*Study 1: measures*

*Study 1: quality indicators*

The QRIS has six indicators of quality: (1) regulatory compliance, (2) teacher–child relationships, (3) learning environment, (4) inclusion of children with special needs, (5) staff qualifications and working conditions, and (6) family and community connections. Table 1 shows the center-based classroom and family child care program level descriptive statistics for the observational measures used to calculate scores for indicator two and three, and program level indicator descriptive statistics are presented in Table 2. Center-based program indicator scores were calculated at the classroom level, and then aggregated to the center level where appropriate. Family child care home QRIS scores are assigned at the program level. Scoring criteria within indicators varies slightly based on the program type and child age (examples of how continuous measures were converted to QRIS ratings are included with each measure, and the full scoring matrix are available in the

Steps to Excellence Program Rating Guides; LA County Office of Child Care, 2009).

*Indicator 1: regulatory compliance*

A licensing review was conducted in collaboration with the state Community Care Licensing Division, and verified for the purpose of QRIS rating. Programs must have had no licensing violations within the past three years to qualify for participation in the QRIS. This item was scored as pass/fail.

*Indicator 2: teacher–child relationships*

Three items were scored to calculate the teacher–child relationship indicator score: (1) teacher–child ratio, (2) group size, and (3) Adult Involvement Scale score (AIS; Howes & Stewart, 1987). Ratio and group size were both observed, and enrollment was also reported by the lead teacher as a second source of information. Enrollment numbers were used to calculate the ratio and group-size item-level score, and were compared with the numbers of observed teachers and children. Observed numbers of children were used if they were higher than the total reported enrollment, and the observed ratio was used if it was higher than reported. One possible option at the low end of the QRIS rating criteria (score of 1) for the teacher–child ratio and group size components for three-to-five year olds was for there to be one lead teacher, one aide with six child development units and a group of 18 children (a ratio of 1:9). At the high end (a score of 5), one lead teacher, two aides, and a maximum group size of 24 was allowed (a ratio of 1:8). Ratio and group size vary at each level for the age group of the majority of the children served in the classroom for center-based programs.

The AIS is a time-sampling rating measuring of the sensitivity and responsivity of interactions provided by the lead teacher. Every 5 min the teacher was observed, and assigned a rating from “ignores” to “intense” on a 7-point scale. The mean level of interactions with children was calculated by weighting the frequency of interactions at each level by the scale point, summing the weighted frequencies, and dividing by the total frequency of observations (psychometric properties reported in Howes & Stewart, 1987). Over the course of a program morning, a minimum of 30 and up to 50

**Table 2**  
QRIS program ratings by domain and element scores.

Indicator	Study1: QRIS descriptive study			Study 2: QRIS simulation study			
	n (programs)	M (%)	SD	n (programs)	M (%)	n (children)	SD
Licensing compliance	254	Pass	–	101	Pass	–	–
T–C relationships	254	3.45	.85	100	4.18	220	.61
Level 1	8	3.15%		0	0.00%	0	
Level 2	22	8.66%		1	1.00%	1	
Level 3	87	34.25%		8	8.00%	20	
Level 4	122	48.03%		63	63.00%	146	
Level 5	15	5.91%		28	28.00%	53	
Learning environment	254	3.90	.79	94	3.76	208	1.27
Level 1	3	1.18%		4	4.26%	12	
Level 2	10	3.94%		15	15.96%	40	
Level 3	45	17.72%		20	21.28%	51	
Level 4	148	58.27%		16	17.02%	33	
Level 5	48	18.90%		39	41.49%	72	
Special needs	254	2.87	1.10	77	1.53	183	.53
Level 1	26	10.24%		37	48.05%	79	
Level 2	78	30.71%		39	50.65%	102	
Level 3	68	26.77%		1	1.30%	2	
Level 4	66	25.98%		–	–	–	
Level 5	16	6.30%		–	–	–	
Staff qualifications & working conditions	249	2.04	.90	84	2.20	193	1.46
Level 1	72	28.92%		36	42.86%	76	
Level 2	115	46.18%		27	32.14%	65	
Level 3	47	18.88%		3	3.57%	6	
Level 4	11	4.42%		4	4.76%	14	
Level 5	4	1.61%		14	16.67%	32	
Family & community	248	3.65	1.36	–	–	–	–
Level 1	19	7.66%					
Level 2	40	16.13%					
Level 3	31	12.50%					
Level 4	66	26.61%					
Level 5	90	36.29%					
Overall score	254	2.76	.83	101	2.59	223	.82
Level 1	14	5.51%		0	0.00%	0	
Level 2	77	30.31%		58	57.43%	127	
Level 3	116	45.67%		31	30.69%	76	
Level 4	45	17.72%		7	6.93%	13	
Level 5	1	0.39%		5	4.95%	7	

ratings are collected for each teacher, and the average AIS code across the observation was used to assign the AIS item score. The 7-point AIS scale was recoded into the 5-point QRIS rating with the following thresholds for the AIS score: (1) AIS < 3, (2) 3.0–3.9, (3) 4.0–4.9, (4) 5.0–5.9, and (5) AIS ≥ 6. The teacher–child relationships indicator score was a weighted average of the three item scores: the AIS item is assigned a .5 weight, the ratio item is assigned a weight of .333, and group size is weighted .167.

*Indicator 3: learning environment*

There are three versions of the environmental rating scales used in the QRIS ratings. The center-based preschool classroom version is the ECERS-R (Harms et al., 1998), center-based infant and toddler classroom version is the ITERS-R (Harms et al., 2000), and the family child care home version is the FCCERS-R (Harms et al., 2007). At the most global level, the Environmental Rating Scales describe the physical classroom environment and materials, and the warmth and sensitivity of teacher–child interactions. Drawing from the extant literature on the ECERS-R (Cassidy et al., 2005; Howes et al., 2008; Mashburn et al., 2008; Pianta et al., 2005), we utilized two factors extracted from the ECERS-R scale: Teaching and Interactions, and Provisions for Learning. We also extended this work to the ITERS-R and FCCERS-R, which demonstrated similar high internal consistency to the ECERS-R factors presented in previous research. Factor scores were collected in an attempt to streamline the time dedicated to collecting ERS data, and to provide a random sample of

ERS items that might represent the full scales (Perlman, Zellman, & Le, 2004). Programs were not notified ahead of time what ERS items would be collected, and were asked to prepare for a full ERS review. The *Teaching and Interactions* factor was a composite of several indicators including staff–child interactions, discipline, supervision, encouraging children to communicate, and using language to develop reasoning skills and was used to measure sensitive interactions between teachers and children (ECERS-R factor  $\alpha = .98$ , ITERS-R factor  $\alpha = .99$ , and FCCERS-R factor  $\alpha = .82$ ). The second factor, labeled *Provisions for Learning*, was a composite of indicators such as furnishings, gross motor equipment, art, dramatic play, and nature/science and was used as a measure of materials (ECERS-R factor  $\alpha = .95$ , ITERS-R factor  $\alpha = .99$ , and FCCERS-R factor  $\alpha = .74$ ). The 7-point ERS factor scores were recoded into the 5-point QRIS rating (noted in parentheses) with the following thresholds for the ERS factor scores: (1) factor score < 3, (2) 3.0–3.9, (3) 4.0–4.9, (4) 5.0–5.9, and (5) ERS ≥ 6. A weighted average was used to assign the learning environment indicator score, with the Teaching & Interactions factor being assigned a weight of .667, and Provisions for Learning .333.

*Indicator 4: inclusion of children with special needs*

The ECE programs provided documentation of their policies regarding the inclusion of children with special needs, staff training and education credits to support the development of children with special needs, as well as their practices to conduct, in conjunction

with parents, developmental screens on all participating children. At the high end of the QRIS rating, programs incorporated all of these practices, and at the low end there were no requirements. This indicator was scored with a weighted average. If children with special needs were enrolled at the ECE site, the weights were as follows: identification and item score (weight = .5), special needs training item score (weight = .333), inclusion of children with special needs (weight = .167). If the site did not have children with special needs enrolled, the weights are modified: identification and item score (weight = .6), and special needs training item score (weight = .4).

#### *Indicator 5: staff qualifications & working conditions*

The ECE programs provided documentation of staff qualifications (including evidence of degrees, early childhood units, and state child development permits), staff compensation scales, professional development training, and staff stability (calculated based on the hiring dates of staff). At the low end of the scale were basic state requirements for program directors, lead and assistant teachers for center-based programs and basic health and safety requirements for providers and assistants in family child care programs. The high end of the QRIS ratings corresponded to NAEYC/NAFCC training and education requirements. Additionally, programs at the high end of the scale must have retained 90% or more of their staff for the past three years, offered benefits, utilized a standardized pay scale, and provided release time for professional development opportunities. All three of the indicator 5 items utilize block scoring within the items, and the indicator was scored as a weighted average of the items. Indicator 5 was scored using the following weighting: staff qualifications (weight = .5), benefits and working conditions (weight = .333), and staff stability (weight = .167).

#### *Indicator 6: family and community connections*

The ECE programs provided documentation of strategies (including flyers, photos, parent handbook, etc.) used to: (1) welcome and encourage involvement all enrolled families, (2) foster strong reciprocal relationships with families, (3) build on family strengths and social connections, and (4) facilitate connections between parents and relevant community resources. Scoring for indicator 6 was based on a block model, and offering more opportunities to connect with families and the community in each category detailed above results in a higher QRIS rating.

#### *Overall QRIS score*

The overall QRIS score was calculated as an average score across the six indicators. However, the overall score was restricted so that the overall score could not be more than one-point higher than the lowest indicator score.

#### *Study 1: procedures*

The Adult Involvement Scale (AIS; Howes & Stewart, 1987) was collected using a time sampling method along with observations of group size and teacher–child ratio. Observations were made in 5-min cycles, with 4 min for observation and 1 min for coding the AIS, resulting in an average of 40 ratings per classroom. Ratings for the ERS measure were cumulative throughout the observation, and a rating was assigned for each item at the conclusion of the observation period.

In centers with multiple classrooms, 50% of classrooms were randomly selected to be observed, prioritizing infant/toddler classrooms, and classrooms serving children with special needs at the request of the local planning agency in an effort to collect data that could provide information relevant to the entire QRIS scoring matrix for programs that served children of differing ages and abilities. Each reviewer observed one classroom at a time for the full

3½ h period. Observations in both studies mostly took place during the morning period, starting at the beginning of the program day. Once the QRIS observation period was completed, one reviewer then completed a review of the required paperwork compiled by programs prior to their site visit. Data from observations and the document review were recorded on data collection forms, with the exception of the staff qualifications paperwork, which was brought back to the university and reviewed by the university-based project coordinator. The data were cleaned, entered into statistical software program, and analyzed at the university. For centers with multiple classrooms, QRIS scores were assigned at the classroom level and aggregated to the center-level.

The observers in both studies were predominantly Spanish-English bilingual, and trained to perform standardized observations in both languages. All data collection staff had extensive field experience in early childhood classroom observation, which was verified by regular reliability field testing, both prior to initiating the site visits and throughout the data collection period. To be certified to conduct a QRIS review, the observer must reach a minimum mean weighted Kappa of less than .70 for each item of the observational measures (AIS & ERS), as well as for the document review. Interrater reliability was re-established every 10th data collection or every three months during active data collection, and inter-rater reliability on all measures was above .9, and remained at that level throughout the three-year pilot study.

#### *Study 1: descriptive study results*

##### *Descriptives of observational measures*

Descriptive statistics from the primary observational measures used to calculate QRIS ratings are presented in Table 1. The median child age of classrooms/programs observed in the QRIS pilot enrolled children in the 36–48 month age category, had an average group size of approximately 15 children, and an adult–child ratio of 1 adult to 4 children across age groups. Average AIS scores were approximately 4.5, and ERS teaching & interaction factor scores were higher on average than provision for learning factor scores.

Additional descriptive analyses were conducted by auspice and child age. Reported and observed ratios and group sizes were similar to one another with observed group sizes and ratios being more favorable than the reported enrollment numbers overall, except in pre-k classrooms where two more children were observed on average ( $n=208$  classrooms;  $M=19.73$ ,  $SD=6.57$ ) than were reported as enrolled in the classroom ( $M=18.16$ ,  $SD=5.46$ ). Family child care homes and infant/toddler classrooms had the smallest group sizes with approximately 6 and 9 children respectively on average ( $F(3,425)=253.38$ ,  $p<.001$ ), and FCCs also had the lowest teacher–child ratios with an average of one adult caring for two to three children ( $F(3,425)=38.12$ ,  $p<.001$ ). Adult involvement scale scores were similar across age groupings (between a score of 4 and 4.5;  $F(3,425)=3.84$ ,  $p>.05$ ), but pre-k classrooms had significantly higher teaching and interaction factor scores on the ERS than FCCs ( $M=6.05$ ,  $SD=0.65$  versus  $M=5.72$ ,  $SD=0.95$ ;  $F(3,425)=4.31$ ,  $p<.01$ ), with a similar pattern for the provisions for learning factor scores ( $M=4.71$ ,  $SD=0.73$  versus  $M=3.78$ ,  $SD=1.05$ ;  $F(3,425)=20.62$ ,  $p<.001$ ).

##### *QRIS scores by program type*

Average QRIS scores were also compared across program type, auspice, and funding sources using Analysis of Variance (ANOVA) to examine differences across the following categories: public preschool centers, Head Start centers, private independently operated centers, and family child care homes with a “large” licensing capacity and a “small” licensing capacity. Small FCCs serve eight or fewer children, and large FCCs may serve up to 14 children. Results of the Analysis of Variance and post hoc analyses are

presented next. When the assumption of homogeneity of variances was violated, results from the Welch statistic were reported ( $F(6,228) = 19.66, p < .001$ ). Head Start programs ( $n = 16$ ) had the highest average QRIS scores ( $M = 3.56, SD = .71$ ) followed by public preschool programs ( $n = 91; M = 3.21, SD = 0.89$ ), centers without Head Start/public pre-k funding ( $n = 14; M = 2.21, SD = 0.63$ ), and FCCs ( $n = 97; M = 2.29, SD = 0.82$ ).

#### QRIS rating descriptive results

QRIS indicator, and overall QRIS program-level scores are presented for the pilot phase of the QRIS in Table 2. Average program scores were in the low-to-middle range, and the smallest numbers of programs occurred at the low and high end of the range. The small number of programs at the low end of the range indicates that most programs were exceeding licensing requirements. Staff qualifications was the quality indicator with the lowest scores, and the family and community connections and the learning environment indicators were the highest on average.

#### QRIS concurrent validity

Correlations between continuous observational measures of child care quality and QRIS scores are presented in Table 1. The direction of the correlations are in the expected direction with positive associations. For example, lower teacher–child ratios are associated with better scores on all of the QRIS indicators. Additionally, the internal consistency of the LA QRIS scoring matrix is in the good range (Cronbach's  $\alpha = 0.71$ ). Supplemental tests were run to assess if the overall scoring reliability of the matrix would be strengthened if any of the domains included in the scoring matrix were removed. Removing any one of the scoring domains did not strengthen the overall internal consistency of the scoring matrix.

### Study 2: QRIS simulation study

#### Study 2: method

##### Study 2: participants

The data used in Study 2 was from a larger longitudinal study of school readiness among low-income children in Los Angeles County, California (for a description of the full study see Fuligni and Howes (2011)). A variety of early childhood education programs serving low-income children was selected to represent a range of diverse learning settings available to low-income children. Since the data collection preceded the QRIS pilot, none of the programs observed were participants in a QRIS. The sample reflects the wide range of early learning programs serving low-income families with preschool-aged children. In the larger study, the sampling procedure involved recruiting programs serving 3-year-old in the first year of the study, as well as, recruiting a comparison group of children not attending a licensed early learning program at age three. Family income was not used as a sampling characteristic, but rather programs and agencies were included if they served low-income families exclusively or made spaces available for families qualifying for subsidies (similar to the recruitment method for ECE programs in the LA QRIS pilot which targeted low-income communities). Children were recruited for the comparison group through several methods that were likely to yield low-income families with similar characteristics, including waiting lists for subsidized child care and targeted flyers in community health care and WIC clinics.

Up to four target children were randomly selected from those eligible in each participating classroom in year 1 of the study, and children who did not enter care until the second year of the study were observed in the classroom or program where they were enrolled during the second year. For this analysis, we included children observed for the first time in a licensed ECE program. The

resulting classrooms in this sample served 3-year-old (60 classrooms/FCCs) and 4-year-old children (41 classrooms/FCCs). One classroom per center-based program is used as a proxy of the overall program score, and the sample includes 40 public preschool classrooms, 34 private preschool classrooms, and 27 family child care homes.

Table 3 presents the descriptive characteristics of the 223 children represented in this largely low-income dataset with an average income-to-needs ratio of less than two-times the federal poverty threshold, and mothers with an average high school level education. Forty-four percent of the children were male. Children were on average 3.7 years old (45 months) when participating in the first assessment. Approximately one-quarter of the children were identified as Spanish-English dual language learners based on their home language at entry into the study. Dual language learners were assessed in both Spanish and English, but only the English assessments are used in this analysis. Most of the children were enrolled in centers (40% in public pre-k and one-third in private centers). Approximately 40% of the sample classrooms or FCCs only enrolled one study child, and the remaining 60 classrooms and FCCs enrolled an average of three study children with a range of 2–5 children observed. The average time between assessments was approximately eight months.

#### Study 2: measures

##### Quality indicators

Quality indicators were calculated as close to the LA QRIS requirements as possible. The following exceptions were made due to differences in the dataset used for the QRIS simulation in Study 2.

##### Indicator 1: regulatory compliance

Due to recruitment criteria, all early care and education programs were assumed to be licensed.

##### Indicator 2: teacher–child relationships

All criteria were followed, except that the AIS observations were focused on individual study children rather than the lead teacher. Therefore, the classroom-level AIS rating represents an average of the teacher–child interaction ratings for children observed in that classroom.

##### Indicator 3: learning environment

The ECERS Academic Environment Scale was used in place of the full ERS in the larger study and includes aspects related to teacher–child interactions and learning materials, so this scale was used as a proxy for the learning environment indicator score. The scale includes a reduced set of ECERS-R and ECERS-E items that was combined to reflect global quality of the academic materials and experiences of children (ECERS Academic Environment Scale). The Early Childhood Environment Rating Scale-Extended (ECERS-E; Sylva, Siraj-Blatchford, & Taggart, 2003) uses the same scoring rubric as the ECERS-R to observationally evaluate the quality of curricular features in various academic domains. This modified approach was used in the larger study, and included some additional items to capture specific content areas which may reflect use of curriculum or program philosophy toward enhancing school-readiness. The scale included an average of five ECERS-R items (22-Blocks, 24-Dramatic Play, 25-Nature/Science, 26-Math/Number, and 20-Art) along with six ECERS-E items (Language items 1–5: Environmental print, Books and literacy, Adult reading with children, Sounds in words, Emergent writing/mark making, and Math item 1: Counting). Each of the items had possible scores ranging from 1 to 7 points. The Academic Environment scale had a mean of 5.21 ( $SD = 1.04$ ) reflecting good quality, and

**Table 3**  
Study 2: QRIS simulation study: child, family, and ECE program characteristics.

	<i>n</i>	<i>M</i> or %	<i>SD</i>	<i>n</i>	<i>M</i> or %	<i>SD</i>
<b>Child-level descriptives</b>						
<i>Child characteristics</i>						
Male	99	44.40%				
<i>Race/ethnicity</i>						
Caucasian	26	12.56%				
Latino	143	69.08%				
African-American	27	13.04%				
Other	11	5.31%				
Child age in months at 1st observation	201	44.65	5.67			
Preschool (3 years)	160	71.75%				
Dual language learner	55	25.70%				
Receptive vocabulary	214	94.43	16.79			
Time between assessments (months)	176	7.91	2.89			
<i>Family and household characteristics</i>						
Income-to-needs ratio	205	1.88	2.3			
Maternal education (Years)	194	12.82	3.27			
<i>Child assessment</i>						
			1st observation		2nd observation	
Mathematics – WJ	44	94.86	17.35	45	97.6	12.84
Receptive lang. – PPVT	154	82.57	17.5	184	81.23	18.31
Pencil tap	188	23.84%	26.66%	193	40.06%	30.99%
Story & print	184	2.89	2.01	192	3.73	2.11
Less anxious	197	2.95	0.57	194	2.98	0.51
Regulated	197	3.11	0.56	194	3.23	0.57
<b>Program-level descriptives</b>						
<i>Race/ethnic proportion</i>						
Caucasian	64	12.52%	24.04%			
Latino	53	50.47%	35.47%			
African American	63	19.56%	28.16%			
Other	64	15.34%	24.70%			
<i>Program type</i>						
Public center	40	39.60%				
Private center	34	33.66%				
Family child care	27	26.73%				

good internal consistency ( $\alpha = .73$ ). Scores on this scale were converted to QRIS scores using the same thresholds as described in Study 1.

**Indicator 4: inclusion of children with special needs**

Data from the teacher surveys were aggregated within centers to determine the following items. (1) *Serve children with special needs*: If the teacher listed at least one child with an individualized education plan in their classroom, they were assigned a marker variable for serving children with special needs. (2) *Special needs training*: if the teacher reported having a certification in special education, this was counted as special needs training. (3) *Developmental screening*: If the teacher reported engaging in any of the following activities, they were given credit for implementing a developmental screening: (a) “arrange a conference with parents to share the information and concerns,” (b) “arrange for a local specialist to observe and evaluate,” (c) “document concern on a special report form,” or (d) “monitor and record the child.” All programs were assigned a score of one for indicator 4, and if they provide evidence of serving children with special needs and provided developmental screenings for children they were assigned a score of 2, and if they also had special needs training they were assigned a score of 3. No scores were assigned above a 3 for this indicator due to insufficient data aligned with this indicator.

**Indicator 5: staff qualifications & working conditions**

The score for this domain was calculated based on the lead teacher’s qualifications as additional staff qualifications were not included in the secondary dataset.

**Indicator 6: family and community connections**

No information was available on family and community connections in the secondary dataset, so this indicator was omitted from the Study 2 simulation.

**Overall QRIS score**

The overall average QRIS score is an unweighted average calculated based on indicators one through five for the Study 2 simulation, with a similar restriction to the LA QRIS that the overall rating is restricted to be within one-point of the lowest-rated indicator score.

**External validation measure**

*Classroom Assessment Scoring System (CLASS)–Pre-K version (LaParo, Pianta, Hamre, & Stuhlman, 2001; LaParo, Pianta, & Stuhlman, 2004)*

The CLASS observational measure of instructional quality was used in this study to assess the validity of the QRIS ratings. In this version of the CLASS, nine dimensions are rated on a Likert scale from 1 to 7, with 1 and 2 indicating low levels; 3, 4, and 5 indicating mid-range levels; and 6 and 7 indicating high levels. Data collection in the preschool and prekindergarten years of this study was conducted before the 2004 version of the CLASS was finalized, so the 2001 version was used in this study. Individual items were averaged to create an overall composite score for each classroom observation, and each of the nine dimensions were grouped together to reflect the most recent version of the CLASS measure. The emotional support domain was calculated as an

average of the following dimensions: positive climate, negative climate (reverse-scored), teacher sensitivity, overcontrol and child engagement. The instructional support domain included the following dimensions: concept development and quality of feedback. Finally, the classroom organization domain was calculated as an average of the behavior management, productivity, and instructional learning formats dimensions. Each subscale met the reliability criteria threshold of an alpha of 0.70. The dimensions are rated every 20 min throughout the program morning for approximately 3½ h, and scores are averaged across the observation period. Training on the CLASS measure by trainers from Teachstone, the publisher of the CLASS, included the use of master-coded video clips to provide examples of various levels (high, low, and moderate) of the CLASS dimensions. Requirements for certification on the CLASS were to code five reliability clips independently and score within one point on the master-coded videotapes on 80% of their scores averaged across the segments, and within one point on each dimension for over 50% of the dimensions. This is a standardized set of reliability tests developed by the publishers of the CLASS.

#### Study 2: simulation – child outcomes

Children participated in individually administered assessments of cognitive/academic and socioemotional outcomes at the beginning and end of the academic year, including: English vocabulary development, math reasoning skills, executive function skills, and socio-emotional behaviors (for a description of the original study, see Fuligni & Howes, 2011). The Peabody Picture Vocabulary Test, 3rd edition (PPVT-3; Dunn & Dunn, 1997) is a receptive language test that can be reliably administered to children as young as 2½ years, and is used extensively in studies of young children's development. For each vocabulary item, the child was asked to point to the picture in an array of four pictures which represents the word spoken by the test administrator. Scores were standardized with respect to norms for child age. Math reasoning was assessed using the Woodcock–Johnson Tests of Achievement Applied Problems subtest (Woodcock, McGrew, & Mather, 2001). This assessment incorporates increasingly difficult questions involving reasoning about numbers, beginning with simple counting of pictures of items and advancing to reasoning about quantity and arithmetic transformations. Applied Problems scores were also standardized relative to age norms. Standardized PPVT and Applied Problems scores in the general population have a mean of 100 and a standard deviation of 15 points. Children's executive functioning was measured using the pencil tap task direct assessment from the Preschool Self-Regulation Assessment (Smith-Donald, Raver, Hayes, & Richardson, 2007), which is an adapted version of a standard peg tapping task, using pencils rather than pegs (adapted from Diamond & Taylor, 1996; Luria, 1966). During the pencil tap task, the child was instructed to tap a pencil on the surface of a table or desk once when the assessor tapped their pencil twice and to tap twice when the assessor tapped once. Children were given three practice trials and 16 test trials, and the pencil tap task score was scored as the percent of correct responses out of the 16 total test trials. Pre-literacy skills were assessed using the story and print concepts task designed for the Family and Child Experiences Study (FACES; Westat Project Team, 2003), in which children were read a book and then asked about the content and the mechanics of reading, such as, identifying where the name of the book is written.

#### Assessor ratings of behavior control and affect

At the end of a 45-min assessment battery, assessors rated the emotion, attention, and behavior of children throughout the assessor–child interaction on a series of dimensions, including positive and negative emotionality, frustration, persistence, and self-regulation. These ratings include items from the

Woodcock–Johnson assessment and a subset of items used by Raver and colleagues in their self-regulation battery (Smith-Donald et al., 2007). Principal components factor analysis of these assessor ratings guided our conceptual grouping of the items into three composite measures, two of which are used in this study: Less Anxious, and Regulated (total variance = 60.19%). Both of the composite measures range from a score of one to four.

#### Study 2: procedures

For Study 2, individual child assessments were collected at the beginning and end of the program year for children in preschool ( $n=160$ ), and at the end of the preschool year and end of the pre-k year for children in pre-k ( $n=63$ ) although the average time between assessments was similar for children first observed at the beginning preschool ( $M=7.98$  months,  $SD=2.36$ ) and for children first observed at the end of preschool ( $M=8.29$  months,  $SD=2.88$ ). In each classroom, bilingual data collectors trained to research standards of inter-rater reliability observed on two different days to collect data for the AIS and the CLASS, and children were assessed during the same occasions by different data collectors. The ECERS observations were conducted on a second day 2–4 months later by a different team of observers. All programs, whether publicly- or privately funded centers or family child care homes, were observed using the same set of tools. In addition, teachers completed questionnaires providing information about their own education background and experience.

For the larger study, inter-rater reliability was established by comparing ratings with ratings from experts (“gold standards”) on that measure through live visits for ECERS, and videotapes for the AIS. For the ECERS, data collectors' mean weighted Kappa with the gold standard for each item score was .69 ( $SD=.07$ ). Data collectors assessing children were also required to complete a rigorous training program to learn the complex administration rules for the PPVT, Woodcock–Johnson, and other measures used in the full assessment battery. Training for the assessor behavior ratings included videotape clips and group consensus discussions. Assessors could not begin fieldwork until they passed an assessment with an age-appropriate child with 100% accuracy on all technical aspects of the procedures, and achieved 90% reliability within one point on each of the behavioral ratings with the gold standard trainer on the assessor–ratings items. Child assessment measures were conducted in the child's ECE program.

#### Study 2: results

##### Analytic plan

The results for study 2 follow a similar analytic plan to study 1: (1) program level descriptives of the observational measures used to calculate QRIS scores and differences by child age and auspice, and (2) intercorrelations between QRIS indicators and continuous measures of classroom quality with comparisons between the two datasets, (3) correlations between continuous quality measures and children's school-readiness scores, and finally, (4) results from models predicting child outcomes from QRIS indicator and overall QRIS scores, including an additional quality indicator under consideration by the LA QRIS. OLS regression models were employed to examine links between the LA QRIS ratings and child outcomes. There were some classrooms in which multiple children were enrolled. So, variability in children's performance on the outcome measures within classrooms was examined (intra-class correlations ranged from .03 to .12). Multi-level modeling was not warranted given low intraclass correlations, and a small sample size at the classroom level. Initial models adjusted for child and family-level covariates in all models: pre-test scores, maternal education, average income-to-needs ratio, child gender, if the

**Table 4**  
Study 2: QRIS simulation study: correlations between continuous measures of child care quality and children’s cognitive/academic, and socioemotional outcomes, and multi-level models controlling for pre-test scores, child, family, and program-level covariates with estimated effect sizes (standardized coefficients) for associations between QRIS quality indicator levels and children’s cognitive/academic, and socioemotional outcomes ( $n=223$  children in 101 early care and education programs).

	Cognitive/academic			Social		
	Math	Receptive language	Executive function	Story & print	Less anxious	Regulated
<i>Continuous quality measures</i>						
Teacher–child ratio	-.04	-.02	.21**	.09	-.06	.04
Group size	.10	-.15*	.07	.01	-.10	-.03
Adult involvement scale	.18	.004	.16*	.21**	-.02	.09
ECERS teaching & interactions	-.05	.08	.15*	.11	.14*	.14*
Years of highest teacher education	-.36*	-.06	.10	.06	-.03	.02
CLASS overall score	.02	-.05	.17*	.18*	.05	.09
<i>QRIS indicators</i>						
Teacher–child relationships	-.04	-.07	-.02	-.02	-.01	.05
Learning environment	-.14	-.05	-.004	-.09	.05	.06
Inclusion of children with special needs	.04	-.03	-.004	-.08	.08	.07
Staff qualifications and working conditions	.02	-.001	-.04	.01	.09	.07
Teacher–child interactions	-.03	-.02	.08	.11+	.08	.07
Overall score	.09	-.06	-.01	-.09	.03	.07

+  $p < .10$ .  
\*  $p < .05$ .  
\*\*  $p < .01$ .  
\*\*\*  $p < .001$ .

child was a Spanish-English dual language learner, child ethnicity, and number of months between assessments. Classroom-level covariates included: the year of assessment (preschool/3 to 4-year-olds or pre-k/4 to 5-year-olds), the percentage of the classroom that was Latino, and program type (public center and private center, with FCC as the comparison group). Non-significant covariates were removed from subsequent models, which yielded a smaller set of final covariates: pre-test, dual-language status, and time between assessments. Because the secondary dataset only had one classroom per center, individual classroom-level data served as an indicator of center quality for children enrolled in center-based programs. Numbers shown in the first part of Table 4 are unadjusted correlations between classroom quality measures and children’s school-readiness. The second part of Table 4 displays standardized effect sizes, and were calculated by standardizing each of the child outcomes and predictors ( $M=0$ ,  $SD=1$ ). Models that examined quality indicator and overall QRIS ratings were modeled with the lowest QRIS scores as the reference group which was most often either level 1 or some combination of levels 1, 2, and 3. Since our primary question is to know if higher QRIS ratings predict increased child learning across the school year, the most conservative approach is to model the contrasts between the highest and lowest level programs in terms of child outcomes. QRIS levels were grouped together so that the reference groups were larger than fifty children to increase our power to detect effects, and increase the precision of results with this small-sample at the program level, and medium-sized sample at the child level. For example, using the child outcome scores from the seven children enrolled in the five ECE programs with the highest QRIS overall rating (score of 5) would not provide enough variability in scores to make a valid comparison with children in lower rated programs. Additionally, no children in the study 2 simulation were enrolled in programs that were scored overall at the bottom level (score of 1), necessitating the combination of program effects across various QRIS indicator and overall levels in order to detect effects. The number of children enrolled in programs with the various QRIS indicator and overall scores are listed in Table 2. Results from the regression analyses are presented in Table 4, and the program levels that are compared to yield each effect size are listed in the text.

There were some missing data for child and family characteristics, and measures of quality, and, therefore quality ratings in the Study 2 simulation (see Table 3). Missing data at the child

and classroom levels in the Study 2 simulation were estimated through multiple imputation procedures in Mplus (Muthén & Muthén, 1998–2011) generating 100 imputed data files. Reported results for the child outcome analyses are averaged across the results of analyses calculated conducted on each of the individual imputed data files to provide estimates of the associations between children’s learning and development at the end of the school year after adjusting for pretest scores, child and family characteristics and indicators of quality. An additional quality indicator, interactions, was also modeled as a potential additional quality indicator that the QRIS could consider incorporating into the overall rating.

*Descriptives of observational measures*

Average program-level descriptives for the observational measures used in the Study 2 QRIS simulation are presented in Table 1. Although Study 2 QRIS simulation includes only children between 3 and 6 years, the average child age was similar to the observational study which included a wider age range. Group sizes, teacher–child ratios, and ERS scale scores were slightly less favorable in the Study 2 simulation, but AIS scores were somewhat higher.

Breaking down the descriptive results of the observational measures by child age and auspice, observed teacher–child ratios were higher in the Study 2 simulation for FCCs with the average adult caring for 5 to 6 children, and observed center-based group sizes were similar to both the 3-year-old and pre-k classrooms in Study 2 (approximately 19–20 children per classroom group) also, following the pattern of descriptive results in Study 1. AIS scores were similar across program types and ages, and pre-k classrooms in centers ( $n=39$ ;  $M=5.98$ ,  $SD=1.17$ ) scored approximately half a point higher on a 7-point scale than 3-year-old classrooms in centers on the ERS Teaching & Interaction factor score ( $n=29$ ;  $M=5.12$ ,  $SD=1.42$ ), and FCCs ( $n=22$ ;  $M=5.09$ ,  $SD=1.51$ ).

*QRIS rating descriptive results*

The overall QRIS scores and associated standard deviations, and patterns of high and low scores across the indicators are similar between Study 1 and Study 2. However, there is some variation across indicators, including higher teacher–child relationships indicator scores in Study 2, and lower special needs indicator scores which have a restricted range in the Study 2 simulation (see Table 2). Higher teacher–child relationship scores in Study 2 are at least in part reflective of the difference in program types between

the composition of the descriptive and simulation studies. FCCs tend to have lower teacher–child relationship scores in both studies (Study 2: FCC scores  $M = 2.87$ ,  $SD = 0.85$  versus centers  $M = 3.83$ ,  $SD = 0.60$ ), and they comprise 38% of the sample in Study 1 and only 27% in Study 2.

There was variability in the proportion of programs rated at the higher end of the range of QRIS scores, and grouping of ratings for comparative purposes in the regression models. Most programs were rated at level 4 (63%) or level 5 (28%) for the teacher–child relationships QRIS quality indicator. The learning environment quality indicator model results compared differences in learning for children in programs rated at level 5 (42% of programs) with children enrolled in programs rated at levels 1 and 2 (~20% of programs). The quality indicator measuring inclusion of children with special needs, had a restricted range of program scores with ratings capped at 3 versus 5 as in the other quality indicators, therefore children in programs rated at the highest two levels of this quality indicator (~52%) were compared with children enrolled in programs with the lowest rating (48%). The regression models for the staff qualifications and working conditions quality indicator (indicator 5) compared the outcomes of children enrolled in level 3–5 programs (~25% of programs) with children enrolled in level 1 programs (~43% of programs). The school-readiness outcomes of children enrolled in ECE programs that received the highest overall QRIS ratings (levels 3–5, 43%) were compared with children with in the lowest-rated programs (levels 1–2, 57%). Finally, a potential indicator of quality was added: teacher–child interactions (measured by the CLASS). Most of the programs in this sample scored a 3 or below on this quality indicator (66%), and 1/3 scored a four or five.

#### QRIS scores by program type

Similar to the results of Study 1, center-based classrooms had higher average overall QRIS scores (3-year-old classrooms:  $n = 40$ ,  $M = 2.64$ ,  $SD = 0.70$ ; pre-k classrooms:  $n = 34$ ,  $M = 2.88$ ,  $SD = 1.07$ ) than FCCs ( $n = 27$ ;  $M = 2.15$ ,  $SD = 0.36$ ) by a spread of more than half a point on a 5-point QRIS scale.

#### Internal consistency of QRIS rating

Evidence of intercorrelations among the quality measures in Study 2 was observed, although QRIS ratings in Study 1 were more highly correlated with continuous measures of structural and observed quality than the simplified version used in the Study 2. Fewer significant associations were observed in Study 2 and the magnitude of those associations was smaller than observed in Study 1. Additionally, unlike Study 1, higher scores on the special needs indicator in Study 2 were negatively associated with observational measures of teacher–child interactions.

#### Measures of child care quality and child outcomes

There were significant, small, positive associations between better scores on measures of child care quality and children's performance on measures of school-readiness (Table 4). For example, smaller group sizes were associated with better receptive vocabulary scores, better teacher–child interactions across measures were associated with executive function and pre-literacy skills, and better teaching & interactions were associated with positive socio-emotional outcomes.

#### QRIS ratings and child outcomes

Results for all models comparing outcomes for children in the highest-rated programs across quality indicators and the highest-rated overall QRIS with children in lower-rated programs did not reach statistical levels of significance. See the bottom of Table 4. One marginally significant result was observed related to the association between teacher–child interactions and pre-literacy skills.

## Discussion

QRIS utilize a variety of quality measures of early care and education quality, organize them into program quality indicators, and translate them into quality ratings (Schaack et al., 2012). This study was designed to examine the concurrent validity of a quality rating matrix, and the predictability of quality ratings on children's outcomes. Results from this study indicate that the organization of quality measures implemented to research-quality standards in this study result in valid program quality indicators and ratings that are consistent with established measures of child care quality. Taken together, these findings suggest that quality ratings based on the LA QRIS scoring criteria can provide useful information to parents and ECE providers to help guide enrollment decisions and program improvement efforts, although further refinement to ratings may be necessary to link them to child outcomes.

#### QRIS scoring

As hypothesized, quality indicator ratings can be reliably combined into an overall QRIS program rating which provides some support for an underlying construct of quality measured by QRIS ratings. In the LA QRIS, quality ratings were positively correlated across indicators, and the quality indicators have good internal consistency. Some indicators may be more distally related to program quality as it pertains to supporting children's development, such as relationships with families and connections to community resources, and may not be expected to be as strongly associated with the other measures of program quality. However, our results suggest that removal of any indicator would reduce the internal consistency of the program-level QRIS rating.

The CLASS is a measure that has widely been shown to be related to child outcomes, and was found to be related at least in part to the four simulated QRIS quality indicators lending some support to the concurrent validity of the QRIS scoring (Hamre & Pianta, 2001). There was some evidence of concurrent validity with the CLASS measure in Study 2, although fewer significant relations between the quality indicators and continuous measures of quality. This might be because the simulated QRIS ratings were based on a simplified set of information with reduced variability in the raw measures as compared with the observed QRIS, and a smaller sample. Additionally, programs that are willing to participate in research might also be more likely to participate in a voluntary QRIS where they would need to give permission to have their quality observed. Therefore, the sample of programs that were observed as part of the simulation study might be representative of programs that would voluntarily participate in a QRIS. Additionally, since these programs are drawn from the same community as the programs that participated in the QRIS, the child outcome results should be a reasonable proxy for the actual QRIS.

Scores on continuous measures of classroom quality were similar to scores on quality ratings observed in other states (Hope Denny, Hallam, & Homer, 2012; Tout et al., 2010). The LA County QRIS uses the AIS to assess the responsivity and intensity of children's interactions with their teachers due to its usefulness across auspices and age groups. No difference in the quality of interactions was observed by age in center-based care settings, but children in pre-k center-based classrooms experienced higher quality interactions with their teachers than children in FCC settings. Additionally, both preschool and pre-k aged children had access to programs with higher provisions for learning scores than children of similar ages in family child care homes. Differences were not observed for infants and toddlers in center-based settings and mixed-age groups in family child care.

### Program ratings

The QRIS Descriptive Study replicates prior findings from empirical studies showing that Head Start and public preschool programs outperform other private centers and most centers outperform FCCs in levels of quality (Li-Grining & Coley, 2006; NICHD ECCRN, 2006). Programs with funding in addition to and often in excess of parent fees and program standards, such as Head Start and public preschool, tend to have higher-quality ratings, and therefore have the potential to have greater impacts on children's outcomes (Barnett, 2010). These results are promising for quality improvement efforts because in combination it is likely that quality-based standards, program infrastructure, and funding contributed to higher levels of quality among programs with standards; these strategies are all under consideration by states who are considering how to both assess and improve the quality of programs within their states.

There is also variability in quality at the classroom level within center-based programs based on the age of the children served when compared to mixed-age groupings in family child care settings. For example, as we would expect with younger children, infants and toddlers in center-based settings have the best group size and ratios, but they also experience classrooms with the lowest learning environment scores of any age group in center-based care. Children in FCC programs experience better group sizes, but larger teacher-child ratios. This point is important to consider as QRIS across the country consider the ages of children included in their QRIS, and the auspices in which they are served. If many children in a given state are being served in programs of lower rated quality, it may be worth consideration to draw them into quality improvement efforts like QRIS.

Similar to QRIS ratings in other states, and another study simulating the quality of early care and education in California, LA County QRIS ratings vary and the distribution is somewhat normally distributed, with most programs in the low to middle range of quality, and few in lowest and highest quality ratings (Karoly & Zellman, 2012; Tout et al., 2010). On average, the staff qualifications and working conditions quality indicator tends to be the indicator with the lowest score across programs. Assistant teacher qualifications tend to be lower than lead teachers and directors, which restricts the overall indicator score. Improving the low qualifications of teachers, and particularly assistant teachers might be considered as an additional outcome of the QRIS. The highest indicator scores on average are in the learning environment and teacher-child relationships quality indicators. The distribution of quality ratings are reflective of previous studies of child care quality showing that most early childhood programs provide care that is of low to medium quality (Li-Grining & Coley, 2006; Peisner-Feinberg et al., 2001). These patterns were also evident in the QRIS Study 2 simulation.

### Program ratings and child outcomes

The Study 2 simulation demonstrated that QRIS indicator and overall scores were not predictive of children's cognitive and social-emotional growth or development across the school year at a statistically significant level. The results of this study are similar to the results from another recent simulation study conducted on QRIS systems, which found a lack of association between QRIS ratings and child outcomes (Sabol et al., 2013). The direction of effects were similar for the teacher-child relationships indicator which includes ratio and group size to simulation study conducted by Sabol and colleagues (2013), however, the magnitude of effects was smaller for the interactions quality indicator in the present study. Considering the score-related weighting within the LA QRIS to favor measures of children's experiences with their teachers

with a hybrid scoring system, we would have expected to see less degradation in the size of effects when the quality indicators were combined together. However, when quality scores were translated into quality ratings and combined together to form quality indicator and overall scores the association between quality and child outcomes was diminished.

One potential explanation for the lack of an association with child outcomes was the truncation of scores due to the QRIS scoring rules related to quality improvement. In an effort to give equal weight to quality improvement efforts across quality indicators, the overall program score was restricted to no more than one point higher than the lowest quality indicator score which could potentially reduce a program's overall score even if they scored higher on many of the quality indicators. As such, the lack of association between quality ratings and child outcomes may be influenced by scoring decisions not directly related to the average quality of the program, but related to other goals of the QRIS. States should take care to note scoring rules and related decisions when validating QRIS ratings. Taken together, these results point to the need for future work examining the relation between QRIS scoring configurations, like inclusion of measures and criteria within indicators, cut-points identified within those measures, and combination of quality indicators into overall QRIS scores that predict gains in children's learning and school-readiness.

The LA QRIS, which is now in its third year of post pilot-phase implementation, is currently revising their QRIS rating protocol and scoring configuration. In alignment with statewide RTT-ELC efforts, the Classroom Assessment Scoring System is being phased into the LA QRIS rating in addition to the use of the overall ERS scores in place of the factor scores (Pianta, La Paro, & Hamre, 2008). There is some evidence in the second study presented in this paper that higher CLASS scores may be predictive of children's growth in pre-literacy skills which can be informative to the statewide QRIS development as well. By committing to continual refinement of instrumentation and scoring criteria based on research-based evidence, stakeholders who guide the implementation of QRIS can ensure that families and programs are getting the most conceptually meaningful ratings possible.

Some indicators in QRIS, such as the teacher-child relationships and learning environment indicators in the LA QRIS, do have established links to child outcomes in the research literature (Early et al., 2007). As such, QRIS have been able to borrow instrumentation that has been developed to measure these constructs in the research, although, recent studies show a lack of association between ratings of environmental quality and child outcomes (Sabol & Pianta, in press). The low staff qualifications and working conditions QRIS scores are particularly troubling from the perspective of quality improvement because wages and benefits are often among the highest expenses for early care and education programs, and improvements would potentially cause programs to incur a considerable increases in their operating expenses to increase benefits and wages to a high enough level to be able to hire and maintain more highly qualified teachers or to provide time and support for existing staff to increase their qualifications. Considering the lack of association between the staff qualifications quality indicator and children's outcomes as has been observed in other studies of the effects of staff education and training, the benefit of such investments need to be carefully considered (Early et al., 2007). Additionally, another secondary analysis examining QRIS scores indicated that increases in teacher education levels are unlikely to yield higher quality interactions as measured by the Classroom Assessment Scoring System (Hamre & Pianta, 2001; Karoly & Zellman, 2012). Other factors may need to be considered as an outcome of the staff qualifications quality indicator, such as inclusion for systems building purposes and the quality of the instruction that teachers receive (Hamre et al.,

2012; Schaack et al., 2012). Future work should address the utility of improvements in staff qualifications in comparison to other measures that might be a better proxy of teacher–child effectiveness.

Considering the LA QRIS completed a three-year pilot phase, moved into an implementation phase, and now is undergoing revisions where significant changes are being made in alignment with the broader statewide RTT-ELC QRIS, an examination of the relation between QRIS ratings and children's school-readiness is timely. As such, the results of this study were shared with the local policy agency, and will hopefully be brought before the statewide consortium that is participating in RTT-ELC to help inform the development of the statewide QRIS and implementation of related QRIS activities. For example, the lengthy process of developing the QRIS that was intentionally inclusive of local and state partners served the partial goal of creating face validity and buy-in from practitioners which may prove useful in the implementation phase of RTT-ELC. Additionally, the evidence presented in this study provided an opportunity for the state to consider if the QRIS rating levels were inclusive of changes in quality that we would expect to predict meaningful changes in children's learning, and if there might be other measures of children's experiences in ECE that are more predictive of children's learning which should be considered for inclusion in the QRIS rating, such as, the CLASS. Finally, the focus of this paper was to investigate if one of the goals of QRIS – to improve children's learning and well-being – may be actualized. However, there may be other goals of QRIS, such as systems-building and access to ECE programs among others, which might better be investigated with alternate outcome measures in evaluations of QRIS systems.

### Limitations

The studies presented in this paper attempted to draw from existing data from the same geographic location to simulate child outcomes related to QRIS ratings. Several limitations related to the secondary data analysis in Study 2 QRIS simulation meant that we were not able to examine the associations between all of the quality domains, differences between each of the quality rating levels, and a wider age range and child outcomes. Additionally, the simulated QRIS scores were higher than the observed QRIS, and especially on the teacher–child relationships indicator. The teacher–child relationships quality indicator was measured differently from the LA QRIS as implemented in its pilot phase. Finally, the concurrent validity was not as strong in the simulation as the observed study; therefore, inferential results should be interpreted with caution as the sample of programs that agreed to participate in the research study might be of slightly higher quality than programs that volunteer for a QRIS rating. Reduced funding for early childhood services since the Study 2 simulation data were collected may have also affected the observed QRIS scores, particularly in the areas of ratio and group size which are heavily dependent on program funding. A greater effect might have been observed if we had been able to observe children in the same program across a longer period of time.

### Conclusions

Collaborative efforts between researchers and local child care policy experts have resulted in a quality measure that has good internal consistency, and concurrent validity. However, more data on programs at the highest levels of QRIS ratings are needed in order to fully test the association between the QRIS ratings and child outcomes, and inform quality improvement efforts (Schaack et al., 2012).

### References

- Barnard, W., Etheridge Smith, W., Fiene, R., & Swanson, K. (2006). *Evaluation of Pennsylvania's Keystone STARS quality rating system in child care settings*. Pittsburgh, PA: University of Pittsburgh Office of Child Development and Penn State Capital Area Health and Human Development Institute.
- Barnett, W. S. (2010). Universal and targeted approaches to preschool education in the United States. *International Journal of Child Care and Educational Policy*, 4(1), 1–12.
- Burchinal, M., Kainz, K., & Yaping, C. (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings* (pp. 11–32). Baltimore, MD: Paul H. Brookes.
- Caronongan, P., Kirby, G., Malone, L., & Boller, K. (2011). *Defining and measuring quality: An in-depth study of five child care Quality Rating and Improvement Systems*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research, and Evaluation.
- Cassidy, D. J., Hestenes, L. L., Hedge, A., Hestenes, S., & Mims, S. (2005). Measurement of quality in preschool child care classrooms: An exploratory and confirmatory factor analysis of the early childhood environment rating scale-revised. *Early Childhood Research Quarterly*, 20, 345–360. <http://dx.doi.org/10.1016/j.ecresq.2005.07.005>
- Child Trends. (2010). *Quality Rating and Improvement Systems for early care and education. Early Childhood Highlights*, 1, 1–4.
- Diamond, A., & Taylor, C. (1996). Development of an aspect of executive control: Development of the abilities to remember what I said and to Do as I say, not as I do. *Developmental Psychobiology*, 29, 315–334. [http://dx.doi.org/10.1002/\(SICI\)1098-2302\(199605\)29:4<315::AID-DEV2>3.0.CO;2-T](http://dx.doi.org/10.1002/(SICI)1098-2302(199605)29:4<315::AID-DEV2>3.0.CO;2-T)
- Dunn, L. M., & Dunn, L. M. (1997). *Examiner's manual for the PPVT-III: Peabody picture vocabulary test-third edition*. Circle Pines, MN: American Guidance Services.
- Early, D., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., et al. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development*, 78, 558–580. <http://dx.doi.org/10.1111/j.1467-8624.2007.01014.x>
- Elicker, J., Langill, C. C., Ruprecht, K., & Kwon, K. A. (2007). *Paths to quality: A child care quality rating system for Indiana. What is its scientific basis?* West Lafayette, IN: Purdue University.
- Fulgini, A. S., & Howes, C. (2011). Experiences of low-income dual language learning preschoolers in diverse early learning settings. In C. Howes, J. T. Downer, & R. C. Pianta (Eds.), *Dual language learners in the early childhood classroom* (pp. 19–43). Baltimore, MD: Paul H. Brookes.
- Hamre, B. K., & Pianta, R. C. (2001). Early teacher–child relationships and the trajectory of children's school outcomes through eighth grade. *Child Development*, 72, 625–638. DOI: 0009-3920/2001/7202-0020.
- Hamre, B., Pianta, R., Burchinal, M., Field, S., LoCasale-Crouch, J., Downer, J., et al. (2012). A course on effective teacher–child interactions: Effects on teacher beliefs, knowledge, and observed practice. *American Educational Research Journal*, 49, 88–123. <http://dx.doi.org/10.3102/0002831211434596>
- Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early childhood environment rating scale (revised ed.)*. New York, NY: Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R. M. (2000). *Infant/toddler environment rating scale (revised ed.)*. New York, NY: Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R. M. (2007). *Family child care environment rating scale (revised ed.)*. New York, NY: Teachers College Press.
- Hope Denny, J., Hallam, R., & Homer, K. (2012). A multi-instrument examination of preschool classroom quality and the relationship between program, classroom, and teacher characteristics. *Early Education & Development*, 23, 678–696. <http://dx.doi.org/10.1080/10409289.2011.588041>
- Howes, C. (1997). Children's experiences in center-based child care as a function of teacher background and adult:child ratios. *Merrill-Palmer Quarterly*, 43, 404–425.
- Howes, C., & Stewart, P. (1987). Child's play with adults, toys, and peers: An examination of family and child care influences. *Developmental Psychology*, 23, 423–430. <http://dx.doi.org/10.1037/0012-1649.23.3.423>
- Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., et al. (2008). Ready to learn? Children's pre-academic achievement in pre-Kindergarten programs. *Early Childhood Research Quarterly*, 23(1), 27–50. <http://dx.doi.org/10.1016/j.ecresq.2007.05.002>
- Karoly, L. A., & Zellman, G. L. (2012). *How would programs rate under California's proposed quality rating and improvement system? Evidence from statewide and county data on early care and education program quality*. Santa Monica, CA: RAND Corporation.
- LaParo, K., Pianta, R. C., Hamre, B., & Stuhlman, M. (2001). *Early elementary classroom quality observation system*. Charlottesville, VA: University of Virginia.
- LaParo, K., Pianta, R. C., & Stuhlman, M. (2004). The classroom assessment scoring system: Findings from the prekindergarten year. *The Elementary School Journal*, 104, 409–426. DOI: 0013-5984/2004/10405-0004.
- Li-Grining, C. P., & Coley, R. L. (2006). Child care experiences in low-income communities: Developmental quality and maternal views. *Early Childhood Research Quarterly*, 21(2), 125–141. <http://dx.doi.org/10.1037/0012-1649.43.1.208>
- Los Angeles County Office of Child Care. (2009a). *Steps to excellence program: Quality rating guide for child care centers*. Retrieved from <http://www.ceo.lacounty.gov/ccp/pdf/STEP/STEP%20Child%20Care%20Center%20Matrix%20Handout.11.8.12.pdf>

- Los Angeles County Office of Child Care. (2009b). *Steps to excellence program: Quality rating guide for family child care homes*. Retrieved from <http://www.ceo.lacounty.gov/ccp/pdf/STEP/STEP%20Child%20Care%20FCC%20%20Matrix%20EN%20Handout.07092012.pdf>
- Los Angeles County Policy Roundtable for Child Care. (2006). *Steps to excellence program implementation plan: Launching a child care quality rating system in Los Angeles County*. Los Angeles, CA: County Office of Child Care.
- Luria, A. R. (1966). *The higher cortical functions in man*. New York: Basic Books.
- Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O. A., Bryant, D., et al. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development*, 79(3), 732–749. <http://dx.doi.org/10.1111/j.1467-8624.2008.01154.x>
- Muthén, L. K., & Muthén, B. O. (1998–2011). *Mplus user's guide* (6th ed.). Los Angeles, CA: Author.
- NICHD Early Child Care Research Network. (2002). Child-care structure → Process → Outcome: Direct and indirect effects of child-care quality on young children's development. *Psychological Science*, 13(3), 199–206. <http://dx.doi.org/10.1111/1467-9280.00438>
- NICHD Early Child Care Research Network. (2003). Does quality of child care affect child outcomes at age 4½? *Developmental Psychology*, 39, 451–469. <http://dx.doi.org/10.1037/0012-1649.39.3.451>
- NICHD Early Child Care Research Network. (2006). Child-care effect sizes for the NICHD study of Early Child Care and Youth Development. *American Psychologist*, 61, 99–116. <http://dx.doi.org/10.1037/0003-066X.61.2.99>
- Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., et al. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, 72(5), 1534–1553. <http://dx.doi.org/10.1111/1467-8624.00364>
- Perlman, M., Zellman, G. L., & Le, V. N. (2004). Examining the psychometric properties of the Early Childhood Environment Rating Scale-Revised (ECERS-R). *Early Childhood Research Quarterly*, 19(3), 398–412. <http://dx.doi.org/10.1016/j.ecresq.2004.07.006>
- Phillips, D. A., Mekos, D., Scarr, S., McCartney, K., & Abbott-Shim, M. (2000). *Within and beyond the classroom door: Assessing quality in child care centers*. *Early Childhood Research Quarterly*, 15, 475–496.
- Phillipsen, L. C., Burchinal, M. R., Howes, C., & Cryer, D. (1997). The prediction of process quality from structural features of child care. *Early Childhood Research Quarterly*, 12, 281–303.
- Pianta, R., Howes, C., Burchinal, M., Bryant, D., Clifford, R., Early, D., et al. (2005). Features of pre-kindergarten programs, classrooms, and teachers: Do they predict observed classroom quality and child–teacher interactions? *Applied Developmental Science*, 9(3), 144–159. [http://dx.doi.org/10.1207/s1532480xads0903\\_2](http://dx.doi.org/10.1207/s1532480xads0903_2)
- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system*. Baltimore, MD: Paul H. Brookes.
- Sabol, T. J., & Pianta, R. C. (2014). Do standard measures of preschool quality used in statewide policy predict school readiness. *Education Finance and Policy*, 9(2), 116–164. <http://dx.doi.org/10.1162/EDFP.a.00127>
- Sabol, T. J., Soliday Hong, S. L., Pianta, R. C., & Burchinal, M. R. (2013). Can ratings of Pre-K programs predict children's learning? *Science*, 341, 845–846. <http://dx.doi.org/10.1126/science.1233517>
- Schaack, D., Tarrant, K., Boller, K., & Tout, K. (2012). *Quality Rating and Improvement Systems: Frameworks for early care and education systems change*. In S. L. Kagan, & K. Kauerz (Eds.), *Early childhood systems: Transforming early learning*. New York, NY: Teachers College Press.
- Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academy Press.
- Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly*, 22(2), 173–187. <http://dx.doi.org/10.1016/j.ecresq.2007.01.002>
- Sylva, K., Siraj-Blatchford, I., & Taggart, B. (2003). *Assessing quality in the early years: Early Childhood Environment Rating Scale-Extension (ECERS-E): Four curricular subscales*. Stoke on Trent, England: Trentham Books.
- Thornburg, K. R., Mayfield, W. A., Hawks, J. S., & Fuger, K. L. (2009 October). *The Missouri quality rating system school readiness study*. Columbia, MO: Center for Family Policy & Research.
- Tout, K., Starr, R., Soli, M., Moodie, S., Kirby, G., & Boller, K. (2010). *Compendium of quality rating systems and evaluations*. Washington, DC: Administration for Children and Families, Department of Health and Human Services.
- Westat Project Team. (2003). *FACES: 2003 Story-and-print concepts*. Rockville, MD: Westat, Inc.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III tests of achievement*. Itasca, IL: Riverside.
- Zellman, G. L., & Fiene, R. (2012). *Validation of Quality Rating and Improvement Systems for early care and education and school-age care, (Research-to-Practice Brief OPRE 2012-29)*. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.
- Zellman, G. L., & Perlman, M. (2008). *Child care Quality Rating and Improvement Systems in five pioneer states: Implementation issues and lessons learned*. Santa Monica, CA: Rand Corporation.