

Child Care Quality Rating and Improvement Systems											
Title	Author(s)	Year	Publication Source	Sample Size (N)	Sample Composition	Data Source	Independent Variable(s)	Dependent Variable(s)	Summary of Findings	Limitations to Causal Inference	Exclusion Criteria
The effects of accountability incentives in early childhood education	Bassok, D., Dee, T. S. & Latham, S.	2019	Journal of Policy Analysis and Management	2,950	Program-by-year data; licensed center-based ECE programs in North Carolina participating in QRIS and ERS rating and serve children not yet in school (originally rated between 2007-2009)	North Carolina Department of Health and Human Services	• Programs falling just below ERS baseline rating of 4.5	Quality measures: • Future star ratings • ERS ratings • Staff-child ratios • Teacher education and experience • Space requirements • Enrollment	• Study found that lower initial QRIS ratings were associated with lower student enrollment, especially among programs in high competition areas • Lower initial ratings also predicted increases in ratings over time (in first two years after rating, effects dropped off after three years) • Lower initial ratings also associated with higher global observed classroom quality • No evidence that the intent to treat w/lower star ratings influenced staff education and experience, space requirements, and staff-child ratios		
Impacts of a child care quality rating and improvement system on child care quality	Boller, K., Paulsell, D., Del Grosso, P., Blair, R., Lundquist, E., Kassow, D.Z., Kim, R. & Raikes, A.	2015	Early Childhood Research Quarterly	117	Two communities in Washington State; sample sizes family child care (FCC) 25 treatment, 23 control; centers 7 treatment, 7 control and classrooms/teachers 30 treatment, 25 control	Observations, interviews, questionnaires (Seeds to Success field test)	• Quality improvement (coaching and grants) services in QRIS	• Observed quality (ratios, ERS, Arnett Caregiver Interaction Scale) • QRIS ratings	• Observed quality improved for treatment group overall, there were higher scores for centers in the treatment group (not FCC) for the CIS total score • Ratios in centers were lower in treatment than control groups • QRIS ratings not impacted by the coaching and quality improvement grants	Short follow up period, small sample size	
Testing for quality thresholds and features in early care and education	Burchinal, M., Xue, Y., Auger, A., Tien, H.C., Mashburn, A., Peisner-Feinberg, E., Cavadel, E.W., Zaslow, M. & Tarullo, L.	2016	Monographs of the Society for Research in Child Development	16,142	Smallest sample size in an individual study used for the meta-analysis reported for N. Individual study samples range from 679-2,966. 16,142 children in total combining all samples. Sample of children is primarily low-income.	Head Start Family and Child Experiences Survey (FACES) 2006 and 2009, Head Start Impact Study, National Center for Early Development and Learning, North Carolina Prekindergarten, My Teaching Partner, Preschool Curriculum Evaluation Research, The Follow-up Study of the Early Head Start Research and Evaluation Project	• Quality thresholds Quality measures: • Global • Interactional • Domain-specific	Child outcomes: • Language • Literacy • Math • Social skills	• No evidence of an association between ECERS-R Total and PPVT, evidence of modest, but positive effect size in higher quality classrooms for ECERS-R Interaction score and PPVT receptive language and social competence (but not in lower-quality classrooms or among the other outcomes - WJ-AP math, WJ-LW literacy, behavior problems) • CLASS Emotional support did not provide evidence of thresholds between high/low quality programs for problem behaviors or social skills, CLASS Instructional Support showed evidence was a stronger positive predictor for literacy (WJ L-W) and language (PPVT) in higher quality versus lower-quality classrooms • Some evidence for thresholds emerged for TBRS literacy predicted higher literacy (WJ-LW) and language skills (PPVT) in higher-quality classrooms • Models for CLASS (both scales) and TBRS do not suggest clear evidence for the linearity of trends or specific cut points		Preschool programs. Does not include assessment of child outcomes among infants and toddlers. Only assesses higher vs. lower quality programs (not more detailed quality levels). Not directly related to QRIS, but indirectly related to QRIS development.
Are ratings from Tiered Quality Rating and Improvement Systems valid measures of program quality? A synthesis of validation studies from Race to the Top-Early Learning Challenge States	Fox, L., McCullough, M., Caronongan, P., Herrmann, M. & Rimdzius, T.	2019	US Department of Education; Mathematica Policy Research; Institute of Education Sciences; National Center for Education Evaluation and Regional Assistance	5,948	The sample includes 9 QRIS for states that received round I RTT-ELC grants. Sample sizes vary by states and outcome measures (CA: 1,524-166; DE: 1,012-108; MA: 402-73, MN: 913-57; OH: 325-10; OR: 304-259; RI: 331-32; WA: 412-100; WI: 725-98). Total sample size includes max sample size from all 9 states.	QRIS administrative data, interviews	• QRIS ratings	• Observed quality • Gains in child development	• Across states, higher-rated programs scored higher on measures of quality than lower-rated programs (statistically significant in 7 of 8 states) • The difference between these programs may not be large enough to lead to differences in children's outcomes • Ratings not associated with statistically significant differences in children's outcomes overall • Positive significant effect on comprehension • Only 2 states found statistically significant better scores for children in higher-rated programs in at least one domain: DE (cognition) and WA (motor skills, comprehension) • Each state also had null results in other domains (DE - social-emotional, WA in 5 other domains) • Even when examining outcomes in the highest vs. lowest rated programs within states, there was no evidence that children in higher-rated programs did better than those attending lower-rated ones	Non-experimental designs of validation studies, at times without appropriate controls and early in implementation programs	Study focused on preschool programs

Child Care Quality Rating and Improvement Systems											
Title	Author(s)	Year	Publication Source	Sample Size (N)	Sample Composition	Data Source	Independent Variable(s)	Dependent Variable(s)	Summary of Findings	Limitations to Causal Inference	Exclusion Criteria
Inequities in access to quality early care and education: Associations with funding and community context	Hatfield, B.E., Lower, J.K., Cassidy, D.J. & Faldowski, R.A.	2015	Early Childhood Research Quarterly	6,882	Original sample was nearly all licensed early care and education programs (total of 8,903), pared down due to inclusion criteria. Final sample included 3,725 family child care homes, 3,157 centers; nested in 619 zip codes	Administrative data from North Carolina Division of Child Development, Head Start, and Office of SChool Readiness; observational data from the North Carolina Rated License Assessment Project; survey data from the Decennial Census 200 Summary File 3	<ul style="list-style-type: none"> Program-level characteristics Community-level characteristics 	<ul style="list-style-type: none"> Quality rating score (total points) 	<ul style="list-style-type: none"> For centers community characteristics (residential stability, affluence) predicted significant variance in licensing points, More at Four, subsidy concentration, and Head Start funds associated with higher licensing points (Head Start, DCD funding moderate impact of community affluence), null effects for concentrated immigration For home-based care, less of variance in total licensing points was predicted by communities, residential stability was marginally associated with lower quality, concentrated affluence was associated with higher quality, the main effect for concentrated disadvantage was not significant 	Cross-sectional. Program, not classroom, assessment of quality; no controls for characteristics of children served in programs	Does not disaggregate by age served
The impact of quality rating and improvement systems on families' child care choices and the supply of child care labor	Herbst., C. M.	2018	Labour Economics	14,270	Pooled cross-section of children from two survey years included in study (2005 n = 6785, 2012 n = 7485); subanalyses also presented among mothers by education level (N w/a high school degree or less = 8127 and N w/more than a high school diploma = 5732); separate analyses on mothers' employment uses CPS data and sample is 259,005	National Household Education Survey, Early Childhood Program Participation; March Demographic Supplement to the Current Population Survey; Quarterly Workforce Indicators series in the Longitudinal Employer-Household Dynamics household data	<ul style="list-style-type: none"> Enactment of QRIS (yes/no) 	<ul style="list-style-type: none"> Family child care choices (nonparental care, hours in care, whether family pays for care, expenditures) Maternal employment The supply and compensation of child care labor (hires, separations, turnover, monthly earnings, new hires' monthly earnings) 	<ul style="list-style-type: none"> Implementation of QRIS lead to shift from parental to non-parental care, and more time in care Advantaged families were more likely to use non-parental, formal care (family, center) were only marginally significant Decreased likelihood of paying for care and lower amount paid for households with lower-educational attainment, the reverse was true for higher-income households Maternal employment higher for mothers of young children, concentrated among advantaged mothers (no change in earnings) Increased choice of center-based employment (higher among women with higher educational attainment) Increase in the supply of new child care workers across all levels of education, but this is matched by the departure level Increase in turnover among workers with lower educational attainment and younger workers 		
Comparisons among quality measures in child care settings: Understanding the use of multiple measures in North Carolina's QRIS and their links to social-emotional development in preschool children	Hestens, L.L., Kintner-Duffy, V., Wang, Y.C., La Paro, K., Mims, S.U., Crosby, D., Scott-Little, C. & Cassidy, D.J.	2015	Early Childhood Research Quarterly	422 children 246 teachers	422 pre-school children; 246 teachers (94 toddler, 97 preschool and 55 school-age) from 101 programs in 40 counties in NC	NC QRIS validation study	<ul style="list-style-type: none"> QRIS star rating level 	<ul style="list-style-type: none"> Quality: <ul style="list-style-type: none"> ECERS-R and -E CLASS Emotional Support/Organization/Instructional Support, C5 Positive/Negative Child outcomes: <ul style="list-style-type: none"> Social skills Internalizing and externalizing behavior problems Learning self-efficacy 	<ul style="list-style-type: none"> Different star levels vary in quality (using ECERS-R, ECERS-E, and CLASS ratings) Evidence of higher negative affect in the classroom is associated with more externalizing behavior problems Higher ECERS-E, lower C5 negative were associated with fewer internalizing behavior problems and (unexpectedly) higher CLASS Emotional support were associated with higher internalizing behavior problems Social skills were positively associated with higher CLASS Instructional support, but negatively associated with higher ECERS-E scores Higher scores on ECERS-R and CLASS instructional support were associated with higher levels of learning self-efficacy, but the effects of higher ECERS-E and lower C5 Negative scores were opposite (unexpected) Star ratings did not statistically significantly predict internalizing problems, social skills, or learning self-efficacy Higher star rating levels were associated with fewer externalizing behavior problems 	Point in time measures	Age group focus

Child Care Quality Rating and Improvement Systems											
Title	Author(s)	Year	Publication Source	Sample Size (N)	Sample Composition	Data Source	Independent Variable(s)	Dependent Variable(s)	Summary of Findings	Limitations to Causal Inference	Exclusion Criteria
Quality Rating and Improvement Systems and children's cognitive development	Jeon, L. & Buettner, C. K.	2015	Early Childhood Research Quarterly	313	313 randomly selected children in preschool classrooms from 36 QRIS-participating full-time child care centers	QRIS evaluation study in state; 2006–2010 American Community Survey 5-year estimates data at the census tract; parent questionnaires	• Quality in QRIS (binary indicator for highest)	• Children's cognitive skills (PPVT, PALS-PreK, WJ-III-AP)	<ul style="list-style-type: none"> • Variability in cognitive skills significant related to child care program • Statistically significant direct association between QRIS and children's cognitive skills (children in the highest level QRIS did better than those in the lower 2 levels) • QRIS moderates the association between socioeconomic risk and child cognitive scores 	Cross-sectional	Preschool classrooms - average age of children was about 4.5 years old
Identifying baseline and ceiling thresholds within the Qualistar Early Learning Quality Rating and Improvement System	Le, V.N., Schaack, D.D., & Messan Setodji, C.	2015	Early Childhood Research Quarterly	300	300 3-5 year old children who attended centers for at least 10 hours per week in 49 community child care centers; eligible centers were located near low-performing schools and served a high-needs population (>50% subsidy children)	Qualistar QRIS validation study	<ul style="list-style-type: none"> • Structural (staff training and education, ratios) • Children's cognitive (WJ, PPVT) • Children's social-emotional outcomes (Student-Teacher Relationship Scale, Child Behavior Inventory) 	• ECERS-R	<ul style="list-style-type: none"> • Measured effect on cognitive and social composite mixed (similar results were found for the PPVT, WJ-AP, and WJ-LWI) • Impact on ECE credits and teacher education also mixed across analyses 	Smaller sample size; did not assess how thresholds may vary by age group; implementation focused, no comparison/control group	Study of three and four year olds
Can rating pre-k programs predict children's learning?	Sabol, T., Soliday Hong, S.L., Pianta, R.C., & Burchinal, M.R.	2013	Science	2,419	2,419 children in 673 public pre-K programs	National Center for Early Development and Learning (NCEDL) multistate study of prekindergarten, State-Wide Early Education Programs (SWEEP) Study	• QRIS ratings	• End of year child outcomes (children's learning)	<ul style="list-style-type: none"> • CLASS measure most predictive of children's learning of five individual QRIS indicators • Interactions predicted statistically significant higher scores in math, prereading, language, and social skills for children in higher-rated programs than those in lower rated programs • This was also true for ECERS and math outcomes, but not for other outcomes and ECERS, or any outcomes and staff quality, ratio/group size, and family partnership • Composite QRIS rating not predictive of child outcomes 	Only control for beginning-of-the-year scores; follow-up (pre-post) design	Only pre-k programs and pre-k aged kids
Validation of the Quality Ratings used in Quality Rating and Improvement Systems (QRIS): A synthesis of state studies	Tout, K., Magnuson, K., Lipscomb, S., Karoly, L., Starr, R., Quick, H., Early, D., Epstein, D., Joseph, G., Maxwell, K., Roberts, J., Swanson, C. & Wenner, J.	2017	Child Trends/OPRE	2,790 child care facilities; 6,566 children	The sample includes 10 QRIS validation studies. Within that, the samples vary. For example the center population in these studies ranges from 71 to 774, family child care range from 24 to 159 (if included), preschoolers range from 332 to 1,611 (if included), and only 2 studies included infants/toddlers (samples 190 and 239).	QRIS validation study, administrative data	• QRIS ratings	• Observed quality • Gains in child development	<ul style="list-style-type: none"> • All 9 states examining observed quality (excluding WA) found at least 1 positive association with observed quality • Among 3 states looking at CLASS Toddler, DE found null results, OR found mixed results, and RI found positive results • MA used the ITERS-R and found positive, statistically significant results • In general, higher ratings had higher quality scores, but the differences between levels was not always statistically significant • Most results on child outcomes were null • Overall, the most common positive associations were found between QRIS rating level and social-emotional outcomes (4 of 6 states) and executive function (3 of 6 states) 	Small sample sizes was a limitation in each study; State validation studies not RCTs or QEDs, often lacked controls	Few studies include infants and toddlers

Child Care Quality Rating and Improvement Systems											
Title	Author(s)	Year	Publication Source	Sample Size (N)	Sample Composition	Data Source	Independent Variable(s)	Dependent Variable(s)	Summary of Findings	Limitations to Causal Inference	Exclusion Criteria
Associations among tiered quality rating and improvement system supports and quality improvement	Yazejian, N., & Iruka, I.U.	2015	Early Childhood Research Quarterly	412	ECE programs in Quality Counts (Miami-Dade County TQRIS) with a Time 2 rating; 342 center-based programs, 70 family child care homes	Miami-Dade TQRIS information management system	<ul style="list-style-type: none"> Quality improvement supports (financial incentives - grants and awards, scholarships for staff, hours of on-site technical assistance (TA)) 	<ul style="list-style-type: none"> Quality rating (total points, stars) Improvement over time 	<ul style="list-style-type: none"> Models accounted for around half of variance for centers Duration and scholarships were positively, significantly associated with an increase in total points Improvement awards associated with lower total points For home-based care, models accounts for even more of variation Main effect was duration 	Participation in TQRIS is voluntary (approximately 40% participation, lower among FCC); correlational study, not causal	Did not disaggregate by ages served in ECE programs
Boosting family child care success in Quality Rating and Improvement Systems	Hallam, R. A., Hooper, A., Buell, M., Ziegler, M., & Han, M.	2019	Early Childhood Research Quarterly	278	278 licensed family child care (FCC) providers and large FCC programs in Delaware that participated in QRIS. n = 100 treatment group, n = 178 comparison group	Delaware administrative data (monthly)	<ul style="list-style-type: none"> Stars Plus participation (vs. regular QRIS supports) 	<ul style="list-style-type: none"> QRIS rating QRIS trajectories 	<ul style="list-style-type: none"> Stars Plus associated with progression through Star Levels The authors also looked at a subsample of providers with education data and no significant differences were found 	No randomization (convenience sample), questions on comparison group appropriateness, geography limits broad generalization	
Effects of Quality Improvement System for child care centers	Ma, X., Shen, J., Kavanaugh, A., Lu, X., Brandi, K., Goodman, J., Till, L., & Watson, G.	2011	Journal of Research in Childhood Education	13	13 centers with valid data (of 47 centers: 24 had valid ITERS baseline data, and 7 had valid ITERS formal data, 4 had data from both assessments; 20 centers had valid ITERS-R baseline data, 34 centers had valid ITERS-R formal data, 9 had data from both assessments)	administrative data	<ul style="list-style-type: none"> QRIS participation year of entry into QRIS 	<ul style="list-style-type: none"> ITERS/ECERS scores (both original and revised versions used) 	<ul style="list-style-type: none"> Study found some changes across all quality indicators – centers demonstrated statistically significant improvements in activities and in overall measure of quality over 13 months (baseline assessment to formal assessment), both strong effect sizes, no other quality indicators statistically significant Centers significantly improved quality of programs across all quality indicators over 13 months, all with large effect sizes The largest improvement was in overall measure of quality Centers entering later were associated with larger improvements in personal care routine, interaction, and program structure 	Pre-post w/o comparison (only two time points, short time frame), very small sample for infant/toddler center population	