

Systemwide Quality Improvement in Early Childhood Education: Evidence From Louisiana

Daphna Bassok

University of Virginia

Preston Magouirk

Common App

Anna J. Markowitz

University of California, Los Angeles

Despite substantial federal, state, and local investments in improving early care and education (ECE), we know little about whether ECE program quality has improved over time. The lack of data tracking the quality of publicly funded ECE programs at scale creates a substantial evidence gap for policymakers attempting to weigh the returns on, and future of, quality improvement policies. Data from Quality Rating and Improvement Systems (QRIS) provide a promising opportunity to address this problem. Using 4 years of data from a mandatory, statewide QRIS covering subsidized child care, Head Start, and state pre-kindergarten, we document systemwide quality and improvement trends over a period of targeted investment in quality improvement statewide. We find improvements in quality overall, across sectors and communities over this period. Results also reveal differential growth across sectors such that quality gaps diminished. This study highlights the potential of statewide ECE data for informing quality improvement efforts.

Keywords: *early childhood education, accountability, QRIS, teacher-child interactions, school readiness*

CHILDREN'S early learning experiences are critical for their development (Shonkoff & Phillips, 2000), and decades of evidence suggests that high-quality early care and education (ECE) can positively affect both short- and long-term outcomes (Phillips et al., 2017; U.S. Department of Health and Human Services & Administration for Children and Families, 2010; Yoshikawa et al., 2013). Unfortunately, access to high-quality ECE remains scarce (Markowitz et al., 2018). Although a large body of research demonstrates the importance of warm and engaging teacher-child interactions in supporting young children's development (Bronfenbrenner & Morris, 2006; Hamre, 2014; Johnson et al., 2016; Weiland et al., 2013; Zaslow et al., 2016), the quality of these interactions varies substantially across ECE programs and is often quite low, particularly in programs serving the most disadvantaged children (Bassok & Galdo, 2015; Votruba-Drzal et al., 2010).

Policymakers have sought to improve ECE program quality in a number of ways. Traditionally the two primary approaches have been increasing minimum operating requirements (e.g., requiring higher teacher education levels in Head Start, tightening regulations in the Child Care and Development Block Grant [CCDBG]) and elevating funding levels (e.g., earmarking public funding for state pre-kindergarten [pre-K]

programs, the 2018 increase in CCDBG funding). More recently, early childhood accountability systems have emerged as a third way. Nearly all states now have Quality Rating and Improvement Systems (QRIS), which are accountability systems that unify standards, evaluate and publicize quality, and typically provide monetary incentives for improvement (Bipartisan Policy Center, 2018; First Five Years Fund, 2019; Head Start Act, 2007; Kaplan & Mead, 2017; Office of Child Care, 2019b; Regenstein & Lipper, 2013; U.S. Department of Education, 2016; Warner-Richter, 2016). All three approaches to improving ECE quality have expanded significantly over the past two decades as policymakers work to increase ECE quality at scale. Despite these sizable public investments, however, we know surprisingly little about ECE quality at the state or system level, and even less about whether quality has improved in response to these investments.

This article highlights the utility of systemwide quality data for answering pressing research and policy questions. Using data from Louisiana's mandatory QRIS, it provides the first longitudinal and systemwide look at ECE quality using a unique panel of *all* publicly funded, classroom-based settings in a state over a 4-year period characterized by substantial investments in ECE quality. It then examines whether and



how quality improvement trends vary across sectors and communities. Quality gaps between private, center-based child care programs and Head Start and pre-K programs are well-documented (Bassok et al., 2016), as are disparities in quality by community advantage (Bassok & Galdo, 2015; Valentino, 2018), but to date, no studies have explored whether the rates of quality improvement also vary across sectors and communities. The study argues that systemwide data are essential for understanding the returns on large-scale quality improvement efforts and demonstrates how systemwide quality data could be used to undergird better research for data-driven policymaking.

Quality in ECE

ECE quality is often thought of as having two components: structural quality and process quality (Burchinal, 2018). Structural measures of quality are distal, regulable factors such as group sizes and teachers' educational credentials, and are hypothesized to support, but not guarantee, high-quality experiences for children (Dowsett et al., 2008; Phillipsen et al., 1997). Process quality refers to children's actual experiences and interactions in their ECE programs. Although policymakers traditionally regulated structural quality features in the hopes of driving improvement, there is mounting evidence that many structural quality measures are not consistently linked to children's learning and development (Early et al., 2007; Friedman-Krauss et al., 2019). In contrast, a large body of literature shows that the quality of interactions children experience in ECE settings shapes their development (e.g., Araujo et al., 2016; Burchinal, 2018; Burchinal et al., 2008; Friedman-Krauss et al., 2019; Hamre, 2014; Hindman & Wasik, 2015; Howes et al., 2008; Vitiello et al., 2018; Weiland et al., 2013; Zaslow et al., 2016).

Unfortunately, studies show that process quality varies considerably across sites and is often too low for children to derive developmental benefits (Burchinal et al., 2010). For instance, national Head Start data show that 75% of classrooms rate as low on a measure of instructional quality and 13% rate as high on a measure of emotional warmth and sensitivity (Aikens et al., 2016).

Variation in ECE Quality Across Sectors

Center-based care in the United States is provided through several sectors, including subsidized child care, federal Head Start, and state-funded pre-K (Whitebook et al., 2014). Although each of these sectors provides publicly funded early education, they have different historic missions and are administered by different agencies, with divergent approaches to defining, measuring, and regulating quality. Existing research indicates that quality indeed varies significantly such that, on average, publicly funded child care centers provide lower-quality ECE experiences than Head Start or pre-K. For example, using national data

from the Early Childhood Longitudinal Study–Birth Cohort, Bassok et al. (2016) found that school-based pre-K and Head Start programs demonstrated lower child-teacher ratios, more frequent reading and mathematics activities, and a higher probability of following written curricula than center-based child care programs. Using the same national data set, another study shows that Early Childhood Environmental Rating Scale (ECERS) scores, which assess both the resources and organization of the learning environment and teacher-child interactions, were higher in Head Start programs than in other center-based programs (Hillemeier et al., 2013). These differences in quality are explained, at least in part, by the sectors' differing goals, funding levels, and approaches to quality improvement, as described below.

Child Care. Federal subsidies for child care were originally funded to support parents' participation in the workforce. Private child care programs can receive public funds by accepting subsidies from the Child Care Development Fund (a combination of the CCDBG and Temporary Aid for Needy Families [TANF] dollars). Families earning less than 85% of the federal poverty line are eligible for a subsidy, but funding is not guaranteed and many families are placed on waitlists.

Public funding for child care programs is lower than for Head Start and pre-K programs (Barnett & Friedman-Krauss, 2016; Friedman-Krauss et al., 2020; Office of Child Care, 2019a). Child care teachers often earn below \$10 per hour and are significantly more likely than teachers in other sectors to report high levels of workplace stress, depend on government assistance programs, and leave their teaching positions (Bassok et al., 2021; Whitebook et al., 1998, 2014). Child care programs are subject to their state's licensing requirements, which set the floor for program quality. Licensing requirements typically place less emphasis on the quality of teacher-child interactions or school readiness than Head Start or pre-K programs and instead focus on safety regulations and features of structural quality (e.g., group size).

There is little systematic evidence about how quality in child care programs—structural or process—has changed over time, although recent initiatives have begun prioritizing and funding quality improvement efforts. For example, the 2014 CCDBG reauthorization required states to set development and training requirements for teachers, implement developmental guidelines from birth to school entry, and spend quality “set-asides” on at least one quality improvement activity (e.g., QRIS, supporting resource and referral agencies; National Conference of State Legislatures, 2015).

Head Start. Introduced in 1965, the goal of the federal Head Start program is to improve early learning and health outcomes for children in families at or below the federal poverty line. Head Start now enrolls over 750,000 three- and four-year-old children per year, as well as over 250,000

infants and toddlers in Early Head Start (Early Childhood Learning and Knowledge Center, 2019a). All Head Start programs are required to meet and report on specific features of program operations annually (e.g., program curricula, student assessment, etc.; Early Childhood Learning and Knowledge Center, 2019b). Since 2007, each grantee, or organization receiving Head Start funds, must also participate in a grant renewal process every 5 years (Administration for Children and Families, 2016; Friedman-Krauss et al., 2019; U.S. Department of Health and Human Services, 2016), which includes an observational assessment of the quality of teacher-child interactions in some program classrooms.

The existing evidence, though limited, suggests that quality has improved in Head Start over time. Data from Head Start's Program Information Reports show a steady increase in structural quality: 72% of lead teachers held at least a bachelor's degree as of financial year 2018 relative to under 25% in 2000 (Bassok, 2013; Early Childhood Learning and Knowledge Center, 2019a). Using multiple waves of nationally representative Head Start data, Aikens et al. (2016) showed large increases between 2006 and 2014 on the ECERS. The same report found that while, on average, classroom warmth and organization did not change between 2009 and 2014, programs were significantly less likely to score in the "low" range (between 1 and 2) on instructional quality, in 2014 than in 2009 (from 85% to 76%).

State-Funded Pre-K. Publicly funded pre-K programs first emerged in Tulsa, Oklahoma, as a way to expand access to high-quality ECE. Programs in 44 states and Washington, D.C., now serve 1.58 million children annually, at a cost of \$5.18 billion (Bartik et al., 2011; Friedman-Krauss et al., 2019; Gormley & Phillips, 2005). Pre-K classrooms often operate within public schools or community organizations and often resemble elementary school classrooms. Teachers are often required to administer school readiness assessments, and many states require lead teachers hold educational credentials similar to those of elementary school teachers in their state (Friedman-Krauss et al., 2019; Pianta & Howes, 2009).

Due to data limitations, efforts to track quality improvement in state pre-K programs have focused on structural quality. Since 2003, the National Institute for Early Education Research (NIEER) has evaluated state pre-K programs on a set of 10 quality benchmarks (e.g., curriculum standards, educational requirements for lead and assistant teachers, the presence of a continuous quality improvement system). No state met all 10 benchmarks in 2003. By 2019, many states made notable progress and four met all 10; still, insufficient funding remains a barrier to improving access and quality in state pre-K (Barnett et al., 2004; Friedman-Krauss et al., 2020).

Although there have been meaningful investments in quality improvement across all three sectors, the limited

and varied data currently available are not ideal for tracking changes in quality *within* a sector and make comparisons of quality improvement *across* sectors infeasible. This study shows how QRIS data could help overcome these limitations.

Variation in ECE Quality Across Communities

Program quality also differs across communities, with lower-quality ECE programs in communities with more low-income, Black, and Hispanic families. For example, one recent article combining data from 11 states showed that children concentrated in high-poverty areas were more likely to experience low-quality care than their peers in lower-poverty areas (Valentino, 2018). Using data from Georgia, Bassok and Galdo (2015) linked data from each of the state's public pre-K programs to local poverty measures and found that ratings of teacher-child interactions were the lowest for programs in the most impoverished communities. Similar work in New York City concluded that public pre-K programs available in predominately Black and Hispanic neighborhoods were rated lower than those available in White neighborhoods (Latham et al., 2019). While quality seems to systematically differ across communities, no available studies have examined whether quality improvement trends differ across communities based on demographic characteristics. This study addresses this limitation.

Quality Data and QRIS

Despite the growing understanding of the importance of quality in early childhood settings, increased awareness that ECE quality varies across sectors and communities, and considerable policy investments in improving quality, there is very little data available to track ECE quality and assess improvement. This data deficit exists in part because of the lack of coordination across publicly funded center-based child care, Head Start, and state pre-K, and their divergent approaches to defining, measuring, and regulating quality. Differences in whether and how each sector measures quality have made it impossible to conduct cross-sector quality comparisons or systemwide analyses, and, as shown above, even within individual sectors, longitudinal data systems tracking quality are rare. Investing in quality improvement without having a way to track the impact of those investments is problematic.

QRIS could provide consistent, longitudinal measures of program quality across sectors. These systems, now operational in over 40 states, typically use the same quality measures across all center-based programs irrespective of sector. Most states' QRIS do not require universal participation across all programs and sectors, however, and the programs that opt to participate likely differ in important ways from those that do not. For this reason, many existing QRIS cannot provide either a complete snapshot of quality across a

state's ECE programs or a way to track quality improvement over time. In contrast, Louisiana's QRIS, which provides the data for the present study, allows for a systemwide, multi-year understanding of program quality across all publicly funded ECE programs, including subsidized child care, Head Start, and state pre-K under the Louisiana Department of Education (LDOE).

Louisiana and Act 3

In 2012, Louisiana passed Act 3, or the Louisiana Early Childhood Education Act (2012), to unify quality standards across all publicly funded ECE programs, increase ECE quality statewide, and ensure all children enter kindergarten prepared. In response, LDOE introduced a universal QRIS and required all publicly funded programs to participate. Based on the large body of research on the importance of teacher-child interactions, Louisiana based programs' QRIS ratings *solely* on the quality of those interactions, as measured by the Classroom Assessment Scoring System (CLASS; Pianta et al., 2008), one of the most widely used measures of teacher-child interactions. CLASS assesses the warmth, organization, and instructional scaffolding in a classroom, and many studies—using both experimental and correlational designs—show that children's growth is larger in classrooms with higher CLASS scores (Araujo et al., 2016; Burchinal et al., 2008; Hindman & Wasik, 2015; Howes et al., 2008; National Institute of Child Health & Human Development Early Child Care Research Network, 2002; Vitiello et al., 2018; Wasik & Hindman, 2011, 2020; Weiland et al., 2013; Williford et al., 2013; Zaslow et al., 2016). While some studies have found only modest or zero correlations between CLASS and child outcomes (e.g., Guerrero-Rosada et al., 2020), there remains widespread agreement that the interactions children experience in early childhood settings are critical, and that improving these interactions is important for creating large-scale quality improvements, motivating LDOE's use of this tool.

Starting in the 2015–2016 school year, the state required two CLASS observations per year in every classroom serving either toddlers (1–2 years) or preschool-age children (3–4 years) within every publicly funded ECE program, including child care, Head Start, and pre-K. These classroom scores were then aggregated up to the program level, and programs received quality ratings based exclusively on these averaged CLASS scores.

In addition to implementing a mandatory QRIS, Louisiana introduced a series of additional efforts that aimed to improve teacher-child interactions (Cannon et al., 2018; Lieberman, 2018). They created a set of tax credits for teachers and programs to incentivize high CLASS scores. They provided targeted supports, including coaching for programs with very low CLASS scores (LDOE, 2020). They also

introduced a free, mandatory teaching credential for child care teachers that emphasizes teacher-child interactions. The focused investment in quality improvement and the tight alignment between these improvement efforts and the way quality is measured in Louisiana make it a particularly promising context to examine systemwide improvement.

Although a key goal of Louisiana's reform was to increase quality across all programs, fostering improvements may have been more challenging in some contexts. For instance, child care programs may have found adapting to the demands of a novel evaluation system particularly difficult. Unlike Head Start and pre-K programs, which had prior exposure to classroom observations of instructional quality, child care programs were not previously evaluated on the quality of teacher-child interactions. At the same time, the introduction of additional resources and supports—as well as increased exposure to the CLASS—may have led to more rapid improvements in child care programs than in Head Start or pre-K. Similarly, patterns of improvement may have varied across Louisiana parishes, based on their demographic characteristics. For instance, those communities with the greatest poverty may have weaker infrastructure for supporting local programs in their efforts to improve the quality of teacher-child interactions. On the other hand, it may be that in the most disadvantaged parishes, the new supports provided following Act 3 resulted in more rapid improvement. The current study explores this potential for variation.

Current Study

Despite substantial federal, state, and local investments in ECE quality improvement, we know little about whether or not programs have improved over time, particularly at scale. This article uses detailed data from Louisiana's mandatory QRIS to describe systemwide changes in program quality over a 4-year period with an intense focus on quality improvement. Specifically, we ask the following questions:

1. What was the initial level of ECE system quality in Louisiana in 2016? How did this vary across sectors and parishes?
2. To what extent did ECE system quality increase in Louisiana between 2016 and 2019? How did improvement trends vary across sectors and parishes?

This study provides the first statewide exploration of ECE quality, quality improvement, and variability by key program and community characteristics and highlights how detailed data collected in the context of a mandatory QRIS could substantially enhance our understanding of quality improvement efforts.

TABLE 1
CLASS Scores and Proficiency by Year and Sector

Program	2016	2017	2018	2019	Change
Statewide (<i>N</i> = 1,871)					
Average score	4.70 (0.71)	4.84 (0.68)	5.00 (0.68)	5.13 (0.59)	+0.43
% Proficient or above	62	70	77	85	+23pp
No. proficient or above	1,025	1,043	1,225	1,301	+276
Total programs	1,640	1,500	1,590	1,522	118
Child care (<i>N</i> = 866)					
Average score	4.30 (0.65)	4.50 (0.63)	4.64 (0.65)	4.87 (0.58)	+0.57
% Proficient or above	40	51	59	73	+33pp
No. proficient or above	294	337	416	472	+178
Head Start (<i>N</i> = 253)					
Average score	4.64 (0.50)	4.72 (0.53)	4.89 (0.46)	5.01 (0.39)	+0.37
% Proficient or above	61	67	80	93	+32pp
No. proficient or above	134	128	157	185	+51
Pre-K (<i>N</i> = 752)					
Average score	5.16 (0.54)	5.21 (0.57)	5.39 (0.53)	5.43 (0.51)	+0.27
% Proficient or above	88	88	94	95	+7pp
No. proficient or above	597	578	652	644	+47

Note. *N* = 1,871 total programs represented. Standard deviations in parentheses. The abbreviation pp stands for percentage points. CLASS = Classroom Assessment Scoring System.

Method

Data

This study uses 4 years of LDOE administrative CLASS data from the 2015–2016 through 2018–2019 school years (we refer to these as 2016, 2017, 2018, and 2019), the 4 years following statewide implementation of Act 3. The underlying data include scores from two CLASS observations per year in every classroom serving toddlers (1- to 2-year-olds) and preschoolers (3- to 4-year-olds) within every publicly funded program in the state. For the current study, we use annual program-level data—that is, scores aggregated across classrooms to create a total program score. All publicly funded programs with observation data for any year (2016–2019) are included (*N* = 1,871). The number of publicly funded programs in Louisiana varied across years, as new programs entered the QRIS and others exited. Table 1 provides the number of publicly funded programs operating each year; Appendix Table A1 provides the overall distribution of programs by sector for each parish across the 4 years of this study.

Measures

Quality. CLASS observations were conducted by local observers who were often elementary school principals, child care or Head Start directors, or other early education professionals who are certified reliable on the CLASS observation protocol. To conduct observations, individuals must

become certified, which requires observers’ ratings of classroom quality to align with those of master CLASS observers across multiple observations.

The CLASS measures the warmth and sensitivity of the adults in the classroom, the supports they provide for language development, and the scaffolding of instructional content for concept development. During the study period, Louisiana used two versions of the CLASS as part of its QRIS. The Toddler CLASS, used in classrooms serving children 15 to 36 months old, scores classrooms in the Emotional and Behavioral Support and Engaged Support for Learning domains. The Pre-K CLASS, used in classrooms where children are 3 to 4 years old, scores classrooms on Instructional Support, Emotional Support, and Classroom Organization domains (Pianta et al., 2008). CLASS scores range from 1 to 7. Within each classroom, all domain-level scores are averaged to generate a classroom-level score. All observation scores are then averaged across classrooms to generate a program score.

LDOE places programs in one of the four categories based on these scores: Programs with scores ranging from 1 to 2.99 are classified as “Unsatisfactory.” Beginning in 2018, programs scoring at this level in two consecutive years were subject to loss of public funding. Programs scoring between 3.0 and 4.49 are classified as “Approaching Proficient”; programs scoring between 4.50 and 5.99 are classified as “Proficient”; and programs scoring 6.00 and above are classified as “Excellent.” In 2019, Louisiana added a new category for programs scoring above 5.50 but below 6.00 named

“High Proficient,” but we use the four categories that were consistent across time.

In this study, we focus on two primary measures of quality considered by Louisiana policymakers: the average program-level CLASS scores and the percentage of programs rated Proficient or above. The two are mechanically related, but many of Louisiana’s policy supports during this period focused on bringing low-performing programs up to proficiency, making it worthwhile to explore shifts on this metric. Because Louisiana’s QRIS focused on program-level ratings, we use those scores in our main analyses, but we replicate our analyses using classroom-level CLASS scores as well. Those results, which are substantively quite similar, are presented in Appendix Tables A2 and A3.

Sector. Louisiana’s publicly funded ECE system includes subsidized child care, Head Start, and school-based pre-K (which includes programs in both public and private schools). Sectors are identified using LDOE administrative data, and of the 1,871 study programs, 866 (46%) are classified as child care programs, 253 (14%) as Head Start programs, and 752 (40%) as pre-K programs. The proportion of programs in each sector varies across parishes (Appendix Table A1).

Child care. In Louisiana, the Child Care Assistance Program (CCAP), which is administered by LDOE, provides CCDBG or TANF subsidies to providers serving low-income children (Louisiana Policy Institute for Children, 2014). As of 2018, over 14,000 children were enrolled in child care using CCAP funds (LDOE, 2019). The average annual subsidy for children was \$3,400 (LDOE, 2018).

Head Start. Nearly 23,000 children in Louisiana are enrolled in Head Start and Early Head Start programs. In 2017, per-child funding in Louisiana was \$7,973 for the 19,000 three- and four-year-olds in Head Start and \$14,685 for more than 3,000 infant to 2-year-olds in Early Head Start (LDOE, 2018).

Pre-K. Louisiana provides pre-K access for 4-year-old children through a variety of programs. First, the Cecil J. Picard LA4 Early Childhood Program (LA4) provides funding for over 16,000 children to attend full-year, school-based pre-K programs across the state. The 8(G) Student Enhancement Block Grant Program or 8(G) is funded by the Louisiana Board of Elementary and Secondary Education and provides funding for about 3,000 low-income, 4-year-old children to attend pre-K programs. Finally, the Nonpublic Schools Early Childhood Development Program (NSECD) funds enrollment for about 1,400 low-income, 4-year-old children in state-approved private schools or child care programs providing 6 hours of daily programming (Louisiana Policy Institute for Children, 2014). Per-child funding for pre-K in Louisiana was \$4,580 in 2017 (LDOE, 2018).

Parish characteristics. We characterize parish risk level using a comprehensive index of local childhood risk that was designed by the Tulane Institute of Early Childhood Mental Health and the Louisiana Department of Health, Office of Public Health, Bureau of Family Health. For each of Louisiana’s 64 parishes, the index combines five economic factors (percentage of unemployed, percentage of births to single mothers, percentage of mothers with less than a high school education, percentage of children under age 5 living below the poverty line, and the median household income as a percentage of the federal poverty level), five health factors (percentage of low birth weight babies, teen birth rate, infant mortality rate, percentage of uninsured children, and maltreatment of children ages 0–5), and two educational factors (preliteracy skills measured at kindergarten entry in 2015 and the percentage of children in publicly funded early childhood programs in 2015; Tulane Institute of Early Childhood Mental Health & Louisiana Department of Health, 2016). We sort parishes into quartiles based on their index scores, with higher index scores demonstrating greater local risk.

Analytic Strategy

This descriptive study highlights the initial level of ECE quality in Louisiana the first year systemwide data were collected and identifies variability in quality across sectors and parishes. We then document changes in quality between 2016 and 2019 and explore whether growth varies across sectors and communities as well. To do so, we calculate average CLASS scores in 2016 statewide, within each sector, and within each quartile of parish risk. We also estimate the proportion of programs that were rated as Proficient or above overall, by sector, and by quartile of parish risk. We then look at changes in quality over the subsequent 3 years, again documenting overall statewide patterns and disaggregating by sector and parish risk. Because our data include the entire population of publicly funded ECE programs in Louisiana, we make simple comparisons of these means, without tests for statistical significance.

Changes in quality over time could be driven by improvements within individual programs and by changes in the composition of programs operating within Louisiana’s QRIS. For instance, even if quality remained stable at all existing programs, the overall quality in the state could improve if low-performing programs are shutting down and/or if higher-quality programs opened after 2016. To isolate the within-program improvement trends between 2016 and 2019 among the 1,458 programs in the sample for more than a single year, we run an econometric program fixed effects model, which estimates an individual intercept and slope for each program and is equivalent to a random effects model in hierarchical linear modeling. This model shows us the average within-program growth over time and assesses whether

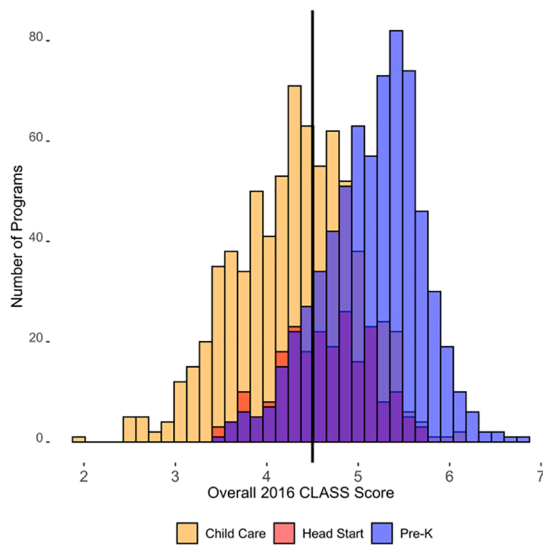


FIGURE 1. *Distributions of 2016 CLASS scores across sectors.*
 Note. Vertical line indicates 4.5-point threshold for proficiency. CLASS = Classroom Assessment Scoring System.

observed changes in CLASS reflect changes to the composition of programs or improvement within programs.

Results

ECE Quality Levels by Sectors and Parish Risk

The first column of Table 1 shows that the statewide average CLASS score in 2016 was 4.70, 0.20 points above the 4.50 threshold for proficiency. In that first year of statewide quality measurement, 62% of programs were rated as Proficient or above. Appendix Table A4 disaggregates the overall scores into individual domains and highlights that initial scores were lowest in the instructional support (pre-K) and engaged support for learning (toddler) domains.

Table 1 also highlights substantial differences in average CLASS scores across sectors. In 2016, the average CLASS score in child care programs was 4.30, 0.20 points below the threshold for proficiency. Forty percent of child care programs were rated Proficient or above. CLASS scores and rates of proficiency were far higher among Head Start (4.64 points or 0.47 of a standard deviation higher, 61% Proficient or above) and pre-K programs (5.16 points or 1.20 standard deviations higher, 88% Proficient or above) than child care, respectively. Figure 1 shows the distribution of program quality by sector for 2016 and indicates that the sector averages mask both considerable variability in scores within each sector and overlap across sectors; many child care and Head Start programs were rated as high or higher than pre-K programs.

Table 2 presents average CLASS scores and rates of proficiency in programs across each of the four quartiles of parish-level risk. Programs in the lowest-risk parishes

(Quartile 1) earned the highest CLASS scores (4.80) and rates of proficiency (65%). Notably, the relationship between parish risk and CLASS was not linear. Average scores and proficiency rates were the lowest in Quartile 3 (4.58 and 58%), rather than in Quartile 4, the highest-risk group. Relative to Quartile 3, average CLASS scores were 0.31 of a standard deviation higher in Quartile 1 (lowest-risk parishes), 0.18 of a standard deviation higher in Quartile 2, and 0.17 higher in Quartile 4 (highest-risk parishes).

Quality Improvement From 2016 to 2019

Table 1 shows that average program-level CLASS scores in Louisiana increased from 4.70 in 2016 to 5.13 in 2019, an increase of 0.61 of a standard deviation. Similarly, the statewide rate of proficiency increased 23 percentage points, from 62% to 85%.

Statewide quality can improve either if individual programs are improving or if there are compositional changes related to quality, for example, if lower-quality programs close. Appendix Table A5 presents results from panel regressions of program-level CLASS scores on indicator variables for year, with and without program fixed effects. We found that the observed increase in CLASS scores over time was due, in large part, to individual program's CLASS scores improving: the average program improved by 0.37 points over the time series (Column 2), and was 20 percentage points more likely to be rated Proficient or above in 2019 than in 2016 (Column 4). On average, the 1,280 programs present across all 4 years of the study improved from 4.79 to 5.17.

We probed this finding further in two ways. First, we explored whether program entry and exit contributed to overall improvement. Appendix Figure A1 shows that programs exiting the sample after the 2016 school year demonstrated lower CLASS scores (4.23) than programs that entered in 2017 (4.47), 2018 (4.73), or 2019 (4.88), suggesting that program composition also played a role in improvement.

Second, we explored variability in growth by initial program quality by estimating average CLASS scores for 2017, 2018, and 2019 for those who did not score Proficient (i.e., Unsatisfactory or Approaching Proficient) in 2016 and those who scored Proficient or above (i.e., Proficient or Excellent). Appendix Figure A2 shows that programs that initially had lower CLASS scores improved more sharply between 2016 and 2019 than those with higher ratings. Specifically, programs that were initially not Proficient increased by 0.87 points on CLASS from 3.96 to 4.83, whereas programs that initially scored at Proficient or above increased by 0.19 points, from 5.14 to 5.33.

Improvement Trends by Sector. As shown in Table 1, growth in CLASS scores was the largest in child care programs,

TABLE 2
CLASS Scores and Proficiency by Year and Parish Risk Quartiles

Program	2016	2017	2018	2019	Change
Statewide (<i>N</i> = 1,871)					
Average score	4.70 (0.71)	4.84 (0.68)	5.00 (0.68)	5.13 (0.59)	+0.43
% Proficient or above	62	70	77	85	+23pp
Quartile 1, Low Risk (<i>N</i> = 381)					
Average score	4.80 (0.76)	5.08 (0.67)	5.19 (0.67)	5.26 (0.64)	+0.46
% Proficient or above	65	80	84	88	+23pp
No. proficient or above	216	243	264	286	+70
Quartile 2 (<i>N</i> = 670)					
Average score	4.71 (0.72)	4.86 (0.67)	4.99 (0.67)	5.12 (0.60)	+0.41
% Proficient or above	65	71	76	85	+23pp
No. proficient or above	388	377	443	466	+78
Quartile 3 (<i>N</i> = 398)					
Average score	4.58 (0.67)	4.71 (0.68)	4.95 (0.65)	5.10 (0.55)	+0.52
% Proficient or above	58	64	77	84	+26pp
No. proficient or above	195	204	256	256	+61
Quartile 4, High Risk (<i>N</i> = 422)					
Average score	4.70 (0.66)	4.71 (0.66)	4.89 (0.69)	5.06 (0.55)	+0.36
% Proficient or above	60	63	73	85	+25pp
No. proficient or above	226	219	262	293	+67

Note. *N* = 1,871 total programs. Standard deviations in parentheses. The abbreviation pp stands for percentage points. CLASS = Classroom Assessment Scoring System.

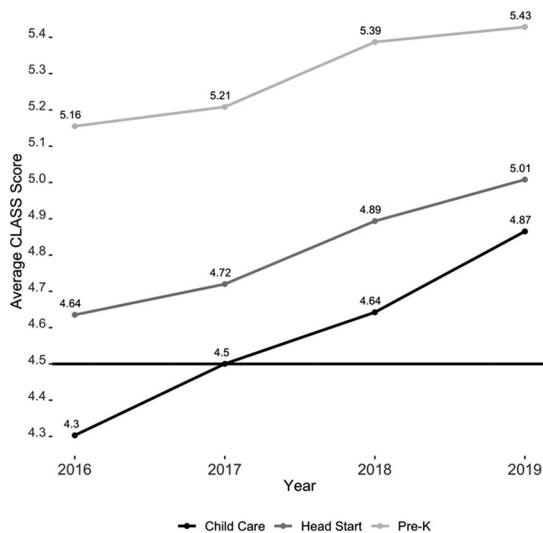


FIGURE 2. *CLASS trends by sector, 2016 to 2019.*
 Note. Horizontal line indicates 4.5-point threshold for proficiency. CLASS = Classroom Assessment Scoring System.

where the average score rose by 0.57 points to 4.87 over the period considered. This is consistent with the evidence above that initially lower-scoring programs showed larger gains in CLASS over the 4-year period. The average score in Head Start programs rose 0.37 points to 5.01. Growth in pre-K program quality was smaller (+0.27 points to 5.43)

than in both child care and Head Start. However, this sector still received the highest scores on average in every year. Figure 2 highlights the steep upward trend in CLASS scores within child care programs and the decreasing size of the gap between average scores for child care and other sectors. The gap between child care and pre-K decreased by more than 33%, and the gap between Head Start and child care was reduced by nearly 60%.

These trends were also reflected in percent proficiency. While the vast majority of programs across sectors earned scores above the 4.50-point threshold for proficiency by 2019, the biggest change in proficiency rates was in the child care sector (Table 1). By the end of the study, 73% of child care programs (up from 40% in 2016), 93% of Head Start programs (from 61%), and 95% of pre-K programs (from 88%) were rated Proficient or above.

Improvement Trends by Parish. Figure 3 presents maps that show average CLASS scores in each parish, from 2016 to 2019. The shading corresponds to quintiles of 2016 CLASS scores. The darkening of the maps from year to year shows that CLASS scores improved in most parishes. By 2019, average CLASS scores in most parishes exceeded 4.82 (the 60th percentile in 2016 scores) and often reached above 5.04 (the 80th percentile in 2016).

Table 2 examines whether rates of improvement in CLASS scores were associated with parish risk scores.

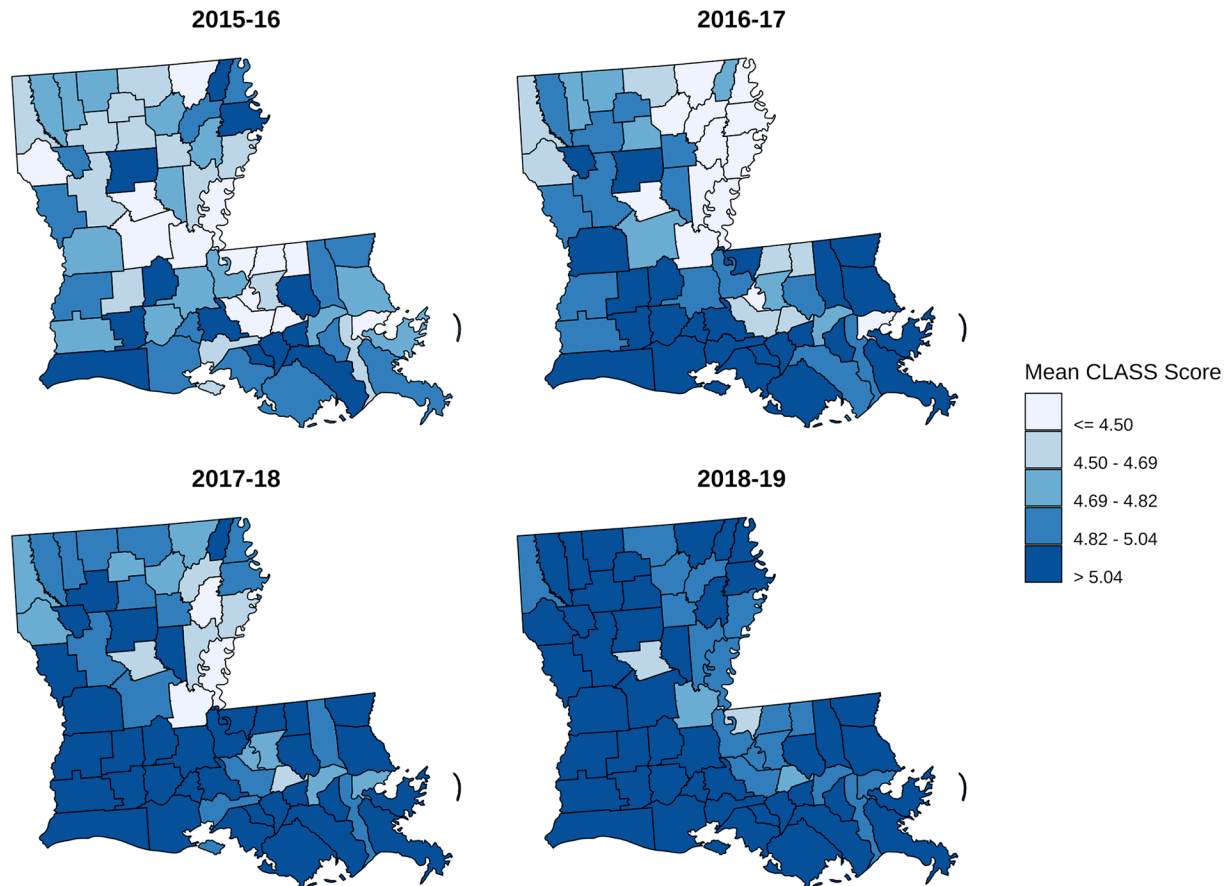


FIGURE 3. CLASS scores across parishes.

Note. In 2015–2016, average CLASS scores at the parish level were divided into quintiles. White represented the lowest quintile and dark blue represented the highest. This same color scheme, using the same 2016 quintiles, is used in all the subsequent years. The dark blue map in 2018–2019 indicates that by 2018–2019 average CLASS scores across nearly all parishes met or exceeded those of the top quintile in 2016. CLASS = Classroom Assessment Scoring System.

Figure 4 shows the growth trajectories by quartiles of parish risk from 2016 to 2019. Overall, initial program growth was slower in higher-risk parishes but trends stabilized over time. Programs in risk Quartiles 1 and 2, the lowest-risk parishes, improved steadily each year, and reached an average score of 5.26 and 5.12 (+0.46 points and +0.41), respectively by 2019. More than 85% of programs in these lower-risk communities were rated Proficient or above in 2019. Quartile 3, which demonstrated the lowest initial CLASS scores, grew more than their counterparts in Quartiles 1, 2, and 4 over the time series (+0.52), consistent with patterns observed across sectors. Like Quartiles 1 and 2, programs in this quartile showed consistent growth each year. In contrast, programs in Quartile 4 demonstrated very little improvement between 2016 and 2017, but afterward showed a more rapid upward trajectory than their more advantaged counterparts. By 2019, programs in the highest risk quartiles looked

similar to programs in Quartile 2, although all three still earned lower ratings and were less likely to be rated Proficient or above than programs in the lowest-risk quartile.

Discussion

Over the past two decades, policymakers have made considerable investments in improving the quality of ECE programs at scale, across cities, states, or nationwide. Unfortunately, there has rarely been systemwide data to measure the quality of ECE programs at a single point in time, let alone whether quality is improving over time. Given the emphasis on and resource commitment to quality improvement, this is a substantial limitation. Investing in quality improvements without a way to assess whether those investments led to change is inefficient. Louisiana’s investment in

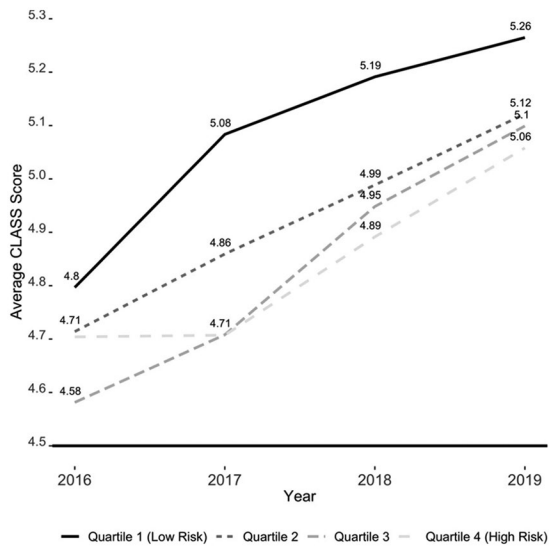


FIGURE 4. *CLASS trends by parish risk quartile, 2016 to 2019.*
Note. Horizontal line indicates 4.5-point threshold for proficiency.
 CLASS = Classroom Assessment Scoring System.

systemwide data collection highlights an alternative. Data collected as part of their QRIS provide rich information on ECE quality both across sectors and over time. They allow for the first study to leverage detailed quality information from *all* publicly funded ECE programs (including subsidized child care, Head Start, and pre-K) statewide to examine quality in both a single year and over time. In doing so, this study both provides a careful look at how quality changed in Louisiana during a period of intensive, targeted investments in ECE quality and highlights how QRIS and other systemwide data collection systems could play an important role in building our understanding of ECE quality improvement efforts at scale.

Contextualizing Program Quality and Improvement in Louisiana

In 2016, average CLASS scores in Louisiana were 4.70, with a standard deviation of 0.71. This level of CLASS was higher, on average, than studies of other ECE contexts, including Georgia’s pre-K program (Bassok & Galdo, 2015) and Head Start (Aikens et al., 2016), but lower than ECE settings known to be of particularly high quality (e.g., Boston public pre-K; Weiland et al., 2013; see Appendix Table A6). The state average masked considerable variability by sector, however, with the average CLASS score in child care more than half a standard deviation lower than the mean, consistent with earlier studies comparing quality across ECE contexts.

Quality also differed by parish risk level, such that parishes categorized as having the lowest risk factors offered higher quality ECE than parishes with higher risk. This finding is similar to Bassok and Galdo’s (2015) study, which

showed that average CLASS domain scores in Georgia’s pre-K programs were between 0.13 and 0.25 points lower in high-poverty ZIP codes than in low-poverty ZIP codes. However, gaps in initial quality were greater across sectors than across parishes. Furthermore, CLASS scores were not the lowest in the most high-risk quartile, but rather in the next quartile up. One possible explanation is that families living in the highest-risk parishes were more likely to attend Head Start and pre-K programs—which target the lowest-income families—than families living in Quartile 3 parishes. In 2016, 47% of programs in Quartile 3 were child care centers compared with 43% in Quartile 4. To the extent that quality is, on average, lower in child care centers, this may account for the observed pattern. In other words, access to quality care in some communities may be driven by sector composition (see Appendix Table A1). Improving access to quality ECE in some parishes may require investments in child care specifically, which may be the main option for many low-income children who just miss the eligibility cutoff for targeted programs, and serves nearly all children aged 0 to 3 years. Though per pupil spending in child care and schools is currently similar (see Measures section above) the cost of educating younger children is higher than for 4- to 5-year-olds largely because of the smaller ratios safe care requires. Moreover, school-based ECE benefits from existing school infrastructure, including administrative support and maintenance staff; shares costs for staffing across a far larger number of students; and is typically not encumbered by specific large costs that child care centers face, including mortgage and rent payments. Focusing additional spending on this sector could free up centers to more boldly pursue a quality improvement agenda.

Quality Improvement

We find that in the 4 years following the introduction of Louisiana’s QRIS, there was substantial quality improvement on CLASS across the state (0.61 *SD*), and that this improvement was driven by within-program changes. Improvement was the largest in child care settings (compared with pre-K and Head Start). Indeed, the proportion of child care programs meeting the proficiency standard increased from 40% to 73% over the time series, narrowing the proficiency gap between child care and pre-K by more than half (from 48 percentage points in 2016 to 22 in 2019). Similarly, we find some evidence of narrowed quality gaps between the high- and low-risk communities. Notably, this growth was nonlinear, such that programs in the most at-risk communities did not show growth until the final 2 years of the study, suggesting that parishes facing the highest levels of community risk may face more substantial barriers to quality improvement and may need targeted supports.

The overall improvements in quality documented in this study are both larger than those observed in other settings

(e.g., Aikens et al., 2016) and likely large enough to be meaningful for children. Due to data limitations, there have not yet been studies that explore whether systemwide improvements in ECE quality lead to developmental gains for children, although many studies do show that children learn more in programs with higher CLASS scores (e.g., Araujo et al., 2016; Burchinal, 2018; Hamre, 2014; Vitiello et al., 2018; Weiland et al., 2013). These findings suggest that the changes in Louisiana led to improved school readiness, and testing this question empirically is a key area for future research, which would be supported by improved data systems like Louisiana's.

Implications for ECE Reform in Other States

Louisiana's efforts to improve ECE quality systemwide may provide useful lessons for other states pursuing quality improvement. Since the passing of Act 3, Louisiana has invested substantially in ECE quality with the goal of ensuring that all programs, irrespective of sector or community, offered high-quality learning opportunities for young children. At the heart of their strategy was their unique QRIS, which was mandated for all publicly funded programs and which focused all sectors on a common definition of quality focused on teacher-child interactions.

LDOE invested heavily in building both buy-in from ECE leadership statewide and an extensive infrastructure to support improvement. For instance, prior to launching the QRIS, LDOE leadership traveled throughout the state to explain the importance of teacher-child interactions, introduce the CLASS, and build excitement about the reform. They conducted two QRIS "pilot years" to ensure a smoother statewide roll out. LDOE subsidized CLASS trainings and professional development opportunities for early educators to build local understanding of and capacity around the CLASS. They have required struggling programs to build improvement plans that center teacher-child interactions. They have incentivized CLASS growth through financial incentives (tax credits), and by publicly celebrating both high-scoring programs and programs that have demonstrated substantial growth.

Notably, these strategies have focused on improvement both at the bottom and at the top of the quality spectrum. Louisiana integrated their QRIS with licensing such that programs that score "Unsatisfactory" in two consecutive years lose their license, creating a floor for quality. They also incentivize continuous improvement through tax credits that increase in size as programs improve their quality rating. Louisiana's recent addition of the "High Proficient" category, for example, demonstrates how the state is creating manageable quality goals and incentivizing centers to reach them. This approach mirrors approaches toward quality improvement used by other states—but because of how Louisiana's QRIS is structured, there is consistency in how

quality is defined at all levels (e.g., at both the bottom and the top), which may make it easier for centers to view quality improvement along a continuum.

Recent survey data from Louisiana suggests that this investment in building buy-in around a single, simple definition of quality has paid off: nearly all ECE program directors and teachers feel that they understand and value the CLASS and that Louisiana's focus on the CLASS—and thereby the QRIS itself—is improving ECE quality (Bassok et al., 2019). This high level of buy-in is encouraging, but ultimately the state wanted to ensure that their various efforts led to actual increases in quality. Their investment in comprehensive data collection—including observation data from every publicly funded ECE classroom in the state—allowed them to track whether quality did in fact increase, and how patterns differed across sectors and communities. In the absence of these data, the state would have no way of knowing, for example, if quality gaps between child care and pre-K had narrowed.

Although our study did not seek to, and cannot, evaluate the causal impact of Louisiana's reform, or of any specific policy they introduced, it does provide compelling evidence that over a period in which the state focused heavily on improving teacher-child interaction quality, and created strong incentives for improvement, CLASS scores went up considerably statewide, across nearly all parishes, and in all sectors. Knowing if and where quality is improving allows policymakers to iteratively refine policies and target resources effectively.

The data the state collected could also facilitate future analyses that rigorously identify the impact of specific policies and initiatives or that explicitly test the returns on particular investment. They provided the data necessary to conduct studies that test the effectiveness of efforts to foster quality improvement at scale.

In this way, Louisiana provides a case study for other cities and states investing in quality improvement. Replication of Louisiana's approach would likely require not just attention to policy rollout, local buy-in, and funding for strategies for quality improvement but also a real investment in statewide data infrastructure. Louisiana highlights the potential power of QRIS or other systemwide data for tracking quality improvement and potentially for evaluating the efficacy of ECE policies. The mandatory nature of their QRIS is particularly important. So long as participation in QRIS is voluntary, or programs can opt to participate only very superficially, it will not be possible to learn from QRIS data in the same way, or to use the data to examine the effects of specific investments on quality systemwide. QRIS could be an essential stepping stone toward building comprehensive ECE data systems that allow us not only to capture key areas of need but also to better understand whether quality is improving and to observe changes that occur in response to policy changes.

Limitations

While our study demonstrates that CLASS scores went up systemwide in Louisiana, it does not explain *why* they did. We cannot identify whether and which policy investments led to changes. We cannot be sure that other factors that may have been changing in Louisiana over time (e.g., changes in the composition or skills of the children in the classroom, changes in the way observers rate classrooms) did not play a role. And we cannot be sure that the changes documented in the current study were meaningful for young children's development.

In Louisiana, and in all states collecting QRIS data, questions remain about whether the data collected capture quality accurately, and whether improvements over time matter for children. Many recent validation studies suggest that QRIS scores fail to accurately capture quality and do not reliably predict child outcomes (e.g. Cannon et al., 2017). The CLASS, which is the sole measure in Louisiana's accountability system, has been widely validated in research contexts and to date remains the ECE quality measure that has most consistently been associated with children's development. However, it is not yet clear whether the changes in CLASS scores documented in this study led to improvements in children's development.

Most studies that have shown a positive relationship between CLASS scores and child outcomes have been based on data collected by research teams in "no-stakes" contexts and have focused on the classroom rather than the program level. In Louisiana, there are three important differences: CLASS is collected locally by practitioners, it is done in a context where increasingly there are meaningful stakes, and it is focused on program-level aggregations of classroom scores.

Louisiana's decision to use local raters to measure quality was purposeful and aimed to create buy-in from local communities. It made implementation of such a large-scale observational system possible; however, there may be tradeoffs with the quality of scores—and thereby our understanding of the quality experienced by children. The reliance on local observers and the incentives attached to program scores, combined with the fact that CLASS scores in Louisiana are higher than those observed in other contexts (see Appendix Table A6) may raise concerns about the validity of the CLASS scores. Specifically, are the scores accurate, and does the growth reported in this study reflect true improvements in quality, or might it capture inflation.

Unfortunately, we do not have data that can directly address this concern: that is, we lack the data to compare the local raters' scores with those of other raters, or to test whether the increases documented on scores correspond to children's developmental gains over a 4-year period. Instead, we raise three related pieces of evidence to bolster confidence about Louisiana's ratings. First, all local CLASS observers are required to be certified on the CLASS using

the standard qualifications for CLASS nationally. Raters are consistently double coded—a process in which a second observer rates simultaneously and scores are compared—to keep scores aligned with national standards. Second, LDOE contracts with certified CLASS observers from outside the local parish to conduct observations in half of Louisiana's ECE classrooms each year and compares those observation scores with those conducted in the same classrooms by local CLASS observers. When there are significant discrepancies, these "third-party" scores replace the local ratings. In the current study, we used the third-party ratings when such a replacement occurred. Finally, previous research in Louisiana has found both a correlation between local raters and independent, research-trained raters ($r = .36$; Vitiello et al., 2018) and a correlation between local observers' CLASS scores and children's learning gains, although this study was conducted prior to the full implementation of consequential accountability.

Future research should continue to address questions of validity in large-scale rollouts of tools initially designed for research (e.g., Head Start; Derrick-Mills et al., 2016). There is a need for developing other tools for measuring process quality in ECE classrooms. Similarly, more work is needed to test whether aggregation to the program level diminishes the validity of CLASS and other classroom-level measures as predictors of children's growth and development (although analyses by classroom suggest that individual classroom quality is also improving, see Appendix Tables A2 and A3).

Conclusion

This article is the first to describe the levels and changes in ECE quality across the universe of publicly funded center-based ECE options in a state. Our results indicate not only that quality increased on average but also that improvement trends reduced disparities in quality across sectors and parishes.

These encouraging results are consistent with the notion that Louisiana's singular focus on measuring and improving teacher-child interactions in ECE settings led to meaningful improvements in young children's experiences in publicly funded ECE settings. Of course, the current analyses do not document the *effect* of Louisiana's reform efforts, nor do they allow for the identification of any causal relationship between specific LDOE improvement initiatives and quality gains. Rather, the goal of the article was to highlight the importance of collecting systemwide data needed to ultimately do those types of analysis, and more broadly to accurately measure the quality of early learning opportunities systemwide. Investing in quality improvement without simultaneously tracking quality will lead to inefficient, potentially wasteful policies, and an inability to adjust course. Conversely, rich data that allow researchers and policymakers to better understand whether and for whom quality is improving is critical.

Appendix

TABLE A1

Sector Distribution of Programs by Parish, 2015–2016 to 2018–2019

Parish	Total Sites	% Child Care	% Head Start	% Pre-K
Acadia	26	38	15	46
Allen	12	25	25	50
Ascension	39	56	5	38
Assumption	6	0	0	100
Avoyelles	23	39	35	26
Beauregard	9	33	11	56
Bienville	10	30	20	50
Bossier	35	37	14	49
Caddo	104	51	12	37
Calcasieu	79	52	3	46
Caldwell	3	33	33	33
Cameron	7	14	29	57
Catahoula/Tensas	6	0	33	67
Central	6	83	0	17
City of Bogalusa	6	67	17	17
Claiborne/Webster	22	41	23	36
Concordia	8	38	25	38
DeSoto	7	29	14	57
East Baton Rouge/City of Baker	175	57	10	33
East Carroll	3	33	33	33
Evangeline	18	28	28	44
Franklin	17	59	6	35
Grant	10	40	20	40
Iberia	33	30	18	52
Iberville	13	54	0	46
Jackson	8	25	38	38
Jefferson	141	50	7	43
Jefferson Davis	11	9	27	64
Lafayette	79	51	19	30
Lafource	29	28	24	48
Lasalle	5	0	20	80
Lincoln	18	61	11	28
Livingston	37	30	5	65
Madison	8	38	25	38
Morehouse	12	42	17	42
Natchitoches	27	52	19	30
Orleans	189	53	12	35
Ouachita/ City of Monroe	81	51	16	33
Plaquemines	8	38	0	62
Pointe Coupee	13	69	0	31
Rapides	80	46	11	42
Red River	4	50	25	25
Richland	11	27	45	27
Sabine	9	11	11	78
St. Bernard	19	63	0	37
St. Charles	18	50	11	39
St. Helena/ East Feliciana	12	25	33	42

(continued)

TABLE A1 (continued)

Parish	Total Sites	% Child Care	% Head Start	% Pre-K
St. James	10	40	0	60
St. John the Baptist	16	50	6	44
St. Landry	50	40	20	40
St. Martin	20	20	50	30
St. Mary	31	23	29	48
St. Tammany	73	63	5	32
Tangipahoa	54	56	17	28
Terrebonne	38	37	13	50
Union	7	29	29	43
Vermilion	22	27	23	50
Vernon	13	23	0	77
Washington	10	40	10	50
West Baton Rouge	8	62	0	38
West Carroll	5	0	20	80
West Feliciana	3	67	0	33
Winn	7	14	14	71
Zachary	8	88	0	12
Statewide	1,871	46	14	40

Note. Counts and percentages reflect all sites with complete data for at least 1 year between 2015–2016 and 2018–2019.

TABLE A2

Classroom-Level CLASS Scores and Proficiency by Year and Sector

Program	2016	2017	2018	2019	Change
Statewide					
Average score	4.73 (0.82)	4.85 (0.77)	5.00 (0.73)	5.12 (0.69)	+0.40
% Proficient or above	63	69	76	83	+20pp
No. proficient or above	3,522	3,666	4,059	4,485	+963
Total Classrooms	5,613	5,350	5,328	5,413	
Child care					
Average score	4.39 (0.79)	4.53 (0.74)	4.72 (0.73)	4.91 (0.72)	+0.51
% Proficient or above	46	54	64	74	+28pp
No. proficient or above	1,168	1,292	1,553	1,782	+614
Head Start					
Average score	4.67 (0.72)	4.76 (0.64)	4.92 (0.63)	5.03 (0.57)	+0.36
% Proficient or above	60	67	75	84	+24pp
No. proficient or above	663	655	687	852	+189
Pre-K					
Average score	5.18 (0.67)	5.27 (0.66)	5.39 (0.59)	5.43 (0.59)	+0.25
% Proficient or above	85	87	92	93	+8pp
No. proficient or above	1,691	1,719	1,819	1,851	+160

Note. Classrooms are drawn from 1,825 total programs. Standard deviations in parentheses. The abbreviation pp stands for percentage points. CLASS = Classroom Assessment Scoring System.

TABLE A3

Classroom-Level CLASS Scores and Proficiency by Year and Parish Risk Quartiles

Program	2016	2017	2018	2019	Change
Statewide					
Average score	4.73 (0.82)	4.85 (0.77)	5.00 (0.73)	5.12 (0.69)	+0.40
% Proficient or above	63	69	76	83	+20pp
No. proficient or above	3,522	3,666	4,059	4,485	+963
Total classrooms	5,613	5,350	5,328	5,413	
Quartile 1					
Average score	4.81 (0.83)	5.05 (0.80)	5.19 (0.74)	5.26 (0.74)	+0.46
% Proficient or above	64	77	82	84	+20pp
No. proficient or above	784	910	951	1,028	+244
Quartile 2					
Average score	4.74 (0.82)	4.87 (0.77)	4.98 (0.74)	5.09 (0.70)	+0.36
% Proficient or above	65	71	75	82	+17pp
No. proficient or above	1,379	1,433	1,517	1,666	+287
Quartile 3					
Average score	4.62 (0.80)	4.72 (0.75)	4.94 (0.72)	5.12 (0.65)	+0.49
% Proficient or above	60	63	76	84	+24pp
No. proficient or above	653	665	824	922	+269
Quartile 4					
Average score	4.71 (0.79)	4.71 (0.71)	4.91 (0.68)	5.02 (0.64)	+0.31
% Proficient or above	61	61	73	82	+21pp
No. proficient or above	706	658	767	869	+163

Note. Classrooms are drawn from 1,825 total programs. Standard deviations in parentheses. The abbreviation pp stands for percentage points. CLASS = Classroom Assessment Scoring System.

TABLE A4

Statewide CLASS Trends, by Domain

Domain	Scores					% Proficient or Above				
	2016	2017	2018	2019	Change	2016	2017	2018	2019	Change
Pre-K Domains										
Instructional Support	3.39 (0.95)	3.52 (0.95)	3.73 (0.95)	3.83 (0.88)	+0.44	13	17	22	23	+10pp
Emotional Support	5.57 (0.69)	5.68 (0.62)	5.81 (0.61)	5.93 (0.54)	+0.36	93	96	97	98	+5pp
Classroom Organization	5.28 (0.78)	5.44 (0.71)	5.57 (0.71)	5.68 (0.64)	+0.40	85	90	93	95	+10pp
Toddler Domains										
ESL	3.30 (0.84)	3.46 (0.85)	3.64 (0.86)	3.92 (0.78)	+0.62	9	12	18	25	+16pp
EBS	5.07 (0.73)	5.28 (0.66)	5.39 (0.69)	5.60 (0.60)	+0.53	80	88	89	96	+16pp

Note. $N = 1,871$ total programs observed across years. Standard deviations are presented in parentheses. “ESL” and “EBS” are abbreviations for the “Engaged Support for Learning” and “Emotional and Behavioral Support” toddler domains, respectively.

TABLE A5

Statewide CLASS Trends and Within-Program Growth by Year

Year	CLASS Score		% Proficient or Above	
Constant (2016)	4.70	4.76	62	66
2017	0.14 (0.02)	0.10 (0.01)	7.03 (1.55)	4.97 (1.29)
2018	0.30 (0.02)	0.26 (0.02)	14.54 (1.52)	12.57 (1.26)
2019	0.43 (0.02)	0.37 (0.02)	22.97 (1.54)	19.67 (1.32)
Program FE		X		X

Note. $N = 1,871$ programs statewide and 1,458 in “Program FE” columns. “Constant” row presents 2016 means. Standard errors in parentheses. Coefficients indicate change relative to 2016. CLASS = Classroom Assessment Scoring System.

TABLE A6
Comparisons of CLASS Scores Across Studies, by Domain

Domain	Louisiana						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Instructional Support	2.06	2.04	4.30	3.26	2.40	3.39	3.83
Emotional Support	5.53	5.49	5.63	—	5.40	5.57	5.93
Classroom Organization	5.17	—	5.10	—	4.80	5.28	5.68
Overall	4.25	—	5.01	—	4.20	4.70	5.13

Note. Scores reflect domain-level averages from several large studies using data from the CLASS. Column (1) presents pre-K data from Georgia (Bassok & Galdo, 2015); Column (2) presents pre-K data from the National Center for Early Development and Learning’s (NCEDL) Multi-State Study of Pre-Kindergarten (Multi-State Study) and the NCEDL-NIEER State-Wide Early Education Programs Study (Burchinal et al., 2010); Column (3) presents pre-K data from Boston (Weiland et al., 2013); Column (4) presents pre-K data from Oklahoma (Johnson et al., 2016). Column (5) presents Head Start data from the 2014 Head Start Family and Child Experiences Survey (FACES) (Aikens et al., 2016); Columns (6) and (7) reflect 2016 and 2019 scores from this study. CLASS = Classroom Assessment Scoring System.

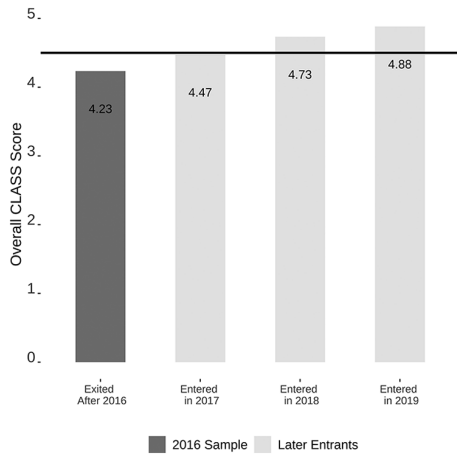


FIGURE A1. Average CLASS scores of exiting programs in 2016, and of entering programs in 2017, 2018, and 2019. Note. Horizontal line indicates 4.5-point threshold for proficiency. CLASS = Classroom Assessment Scoring System.

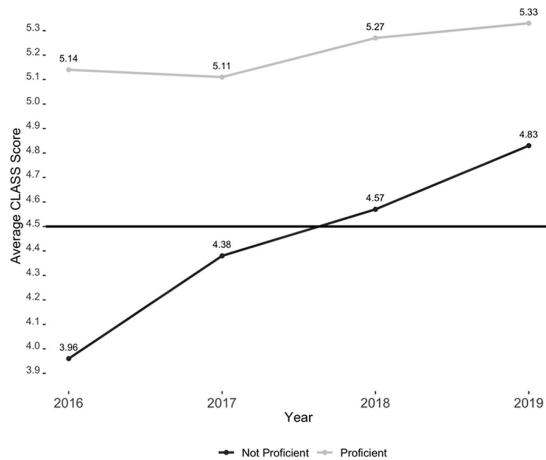


FIGURE A2. Average CLASS scores from 2016 to 2019, by 2016 proficiency status. Note. CLASS = Classroom Assessment Scoring System.

References

Administration for Children and Families. (2016). *Head Start performance standards*. <https://www.federalregister.gov/documents/2016/09/06/2016-19748/head-start-performance-standards>

Aikens, N., Bush, C., Gleason, P., Malone, L., & Tarullo, L. (2016). *Tracking quality in Head Start classrooms: FACES 2006 to FACES 2014* (OPRE Report No. 2016–82). Mathematica Policy Research. https://www.acf.hhs.gov/sites/default/files/opre/faces_tracking_quality_in_head_start_100716_b508.pdf

Araujo, M. C., Carneiro, P., Cruz-Aguayo, Y., & Schady, N. (2016). Teacher quality and learning outcomes in kindergarten. *Quarterly Journal of Economics*, *131*(3), 1415–1453. <https://doi.org/10.1093/qje/qjw016>

Barnett, W. S., & Friedman-Krauss, A. H. (2016). *State(s) of Head Start*. National Institute for Early Education Research. <http://nieer.org/headstart>

Barnett, W. S., Robin, K. B., Hustedt, J. T., & Schulman, K. (2004). *State of preschool: 2003*. National Institute for Early Education Research. <http://nieer.org/state-preschool-yearbooks/state-preschool-2003>

Bartik, T., Gormley, W., & Adelstein, S. (2011). *Earnings benefits of Tulsa’s pre-K program for different income groups*. Upjohn Research. <https://doi.org/10.17848/wp11-176>

Bassok, D. (2013). Raising teacher education levels in Head Start: Exploring programmatic changes between 1999 and 2011. *Early Childhood Research Quarterly*, *28*(4), 831–842. <https://doi.org/10.1016/j.ecresq.2013.07.004>

Bassok, D., Bellows, L., Markowitz, A. J., & Sadowski, K. C. (2021). New evidence on teacher turnover in early childhood. *Educational Evaluation and Policy Analysis*, *43*(1), 172–180. <https://doi.org/10.3102%2F0162373720985340>

Bassok, D., Fitzpatrick, M., Greenberg, E., & Loeb, S. (2016). Within- and between-sector quality differences in early childhood education and care. *Child Development*. <https://doi.org/10.1111/cdev.12551>

Bassok, D., & Galdo, E. (2015). Inequality in preschool quality? Community-level disparities in access to high-quality learning environments. *Early Education and Development*, *27*(1), 128–144. <https://doi.org/10.1080/10409289.2015.1057463>

- Bassok, D., Markowitz, A., Smith, A., & Coffey, M. (2019). *Building buy-in: Teacher and leader views on a statewide effort to improve teacher-child interactions* (Policy Brief). University of Virginia.
- Bipartisan Policy Center. (2018). *Creating an integrated efficient early care and education system to support children and families: A state-by-state analysis*. <https://bipartisanpolicy.org/report/ece-administration-state-by-state/>
- Bronfenbrenner, U., & Morris, P. A. (2006). The bioecological model of human development. In R. M. Lerner, & W. Damon (Eds.), *Handbook of child psychology: Theoretical models of human development* (pp. 793–828). John Wiley.
- Burchinal, M. (2018). Measuring early care and education quality. *Child Development Perspectives*, 12(1), 3–9. <https://doi.org/10.1111/cdep.12260>
- Burchinal, M., Howes, C., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O. (2008). Predicting child outcomes at the end of kindergarten from the quality of pre-kindergarten teacher-child interactions and instruction. *Applied Developmental Science*, 12(3), 140–153. <https://doi.org/10.1080/10888690802199418>
- Burchinal, M., Vandergrift, N., Pianta, R., & Mashburn, A. (2010). Threshold analysis of association between child care quality and child outcomes for low-income children in pre-kindergarten programs. *Early Childhood Research Quarterly*, 25(2), 166–176. <https://doi.org/10.1016/j.ecresq.2009.10.004>
- Cannon, J. S., Meyers, S., & Kaufman, J. H. (2018). *Raising the bar for early childhood education: Early signals on how Louisiana's education policy strategies are working for early childhood providers and community networks*. Rand Corporation. https://www.rand.org/pubs/research_reports/RR2303.html
- Cannon, J. S., Zellman, G. L., Karoly, L. A., & Schwartz, H. L. (2017). *Quality rating and improvement systems for early care and education programs*. Rand Corporation. <https://www.rand.org/pubs/perspectives/PE235.html>
- Derrick-Mills, T., Burchinal, M., Peters, H. E., De Marco, A., Forestieri, N., Fyffe, S., Hanson, D., Heller, C., Pratt, E., Sandstrom, H., Triplett, T., & Woods, T. (2016). *Early implementation of the Head Start designation renewal system: Volume I* (OPRE Report 2016-75a). Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. https://www.urban.org/sites/default/files/2016/11/29/early_implementation_drs_volume_one.pdf
- Dowsett, C. J., Huston, A. C., Imes, A. E., & Gennetian, L. (2008). Structural and process features in three types of child care for children from high and low income families. *Early Childhood Research Quarterly*, 23(1), 69–93. <https://doi.org/10.1016/j.ecresq.2007.06.003>
- Early, D. M., Maxwell, K. L., Burchinal, M., Alva, S., Bender, R. H., Bryant, D., Cai, K., Clifford, R. M., Ebanks, C., Griffin, J. A., Henry, G. T., Howes, C., Iriondo-Perez, J., Jeon, H.-J., Mashburn, A. J., Peisner-Feinberg, E., Pianta, R., Vandergrift, N., & Zill, N. (2007). Teachers' education, classroom quality, and young children's academic skills: Results from seven studies of preschool programs. *Child Development*, 78(2), 558–580. <https://doi.org/10.1111/j.1467-8624.2007.01014.x>
- Early Childhood Learning and Knowledge Center. (2019a, April 2). *Head Start Program Facts*. <https://eclkc.ohs.acf.hhs.gov/about-us/article/head-start-program-facts-fiscal-year-2018>
- Early Childhood Learning and Knowledge Center. (2019b). *Head Start Program Performance Standards*. U.S. Department of Health and Human Services. Administration for Children and Families. <https://eclkc.ohs.acf.hhs.gov/policy/45-cfr-chap-xiii>
- First Five Years Fund. (2019). *Child Care and Development Block Grant (CCDBG)*. <https://www.ffyf.org/issues/ccdbg/>
- Friedman-Krauss, A. H., Barnett, W. S., Garver, K. A., Hodges, K. S., Weisenfeld, G. G., & DiCrecchio, N. (2019). *The State of Preschool 2018: State preschool yearbook*. National Institute for Early Education Research. <http://nieer.org/state-preschool-yearbooks/2018-2>
- Friedman-Krauss, A. H., Barnett, W. S., Garver, K. A., Hodges, K. S., Weisenfeld, G. G., & Gardiner, B. A. (2020). *The State of Preschool 2019: State preschool yearbook*. National Institute for Early Education Research. nieer.org/state-preschool-yearbooks/2019-2
- Gormley, W. T., & Phillips, D. (2005). The effects of universal pre-K in Oklahoma: Research highlights and policy implications. *Policy Studies Journal*, 33(1), 65–82. <https://doi.org/10.1111/j.1541-0072.2005.00092.x>
- Guerrero-Rosada, P., Weiland, C., McCormick, M., Hsueh, J., Sachs, J., Snow, C., & Maier, M. (2020). Null relations between CLASS scores and gains in children's language, math and executive function skills: A replication and extension study. *Early Childhood Research Quarterly*, 54(1), 1–12.
- Hamre, B. (2014). Teachers' daily interactions with children: An essential ingredient in effective early childhood programs. *Child Development Perspectives*, 8(4), 223–230. <https://doi.org/10.1111/cdep.12090>
- Head Start Act, 42 USC 9801 (2007). <https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/hs-act-pl-110-134.pdf>
- Hillemeier, M. M., Morgan, P. L., Farkas, G., & Maczuga, S. A. (2013). Quality disparities in child care for at-risk children: Comparing Head Start and non-Head Start settings. *Maternal and Child Health Journal*, 17(1), 180–188. <https://doi.org/10.1007/s10995-012-0961-7>
- Hindman, A. H., & Wasik, B. A. (2015). Building vocabulary in two languages: An examination of Spanish-speaking dual language learners in Head Start. *Early Childhood Research Quarterly*, 31, 19–33. <https://doi.org/10.1016/j.ecresq.2014.12.006>
- Howes, C., Burchinal, M., Pianta, R. C., Bryant, D., Early, D. M., & Clifford, R. (2008). Ready to learn? Children's preacademic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly*, 23, 27–50.
- Johnson, A. D., Markowitz, A. J., Hill, C. J., & Phillips, D. A. (2016). Variation in impacts of Tulsa pre-K on cognitive development in kindergarten: The role of instructional support. *Developmental Psychology*, 52(12), 2145–2158.
- Kaplan, M., & Mead, S. (2017). *The best teachers for our littlest learners? Lessons from Head Start's Last Decade*. Bellwether Education Partners. https://bellwethereducation.org/sites/default/files/Bellwether_HeadStartWorkforce.pdf
- Latham, S., Jennings, J., Corcoran, S., & Sattin-Bajaj, C. (2019, November 7). *Racial disparities in universal pre-K in New York City* [Conference session]. 2019 APPAM Fall Research Conference, Denver, CO, United States. <https://appam.confex.com/appam/2019/webprogram/Paper32534.html>

- Lieberman, A. (2018). *Lessons from the Bayou State: Three reforms for improving teaching and caregiving*. New America. <https://www.newamerica.org/education-policy/reports/lessons-louisianas-early-childhood-system/>
- Louisiana Department of Education. (2018). *Proposed improvements: Child Care ASSISTANCE Program (CCAP)*. <https://www.louisianabelieves.com/docs/default-source/early-childhood/ecce-ac-7-22-proposed-improvements-to-ccap.pdf?sfvrsn=2>
- Louisiana Department of Education. (2019). *Early Childhood Care and Education: 2018 annual report*. https://www.louisianabelieves.com/docs/default-source/early-childhood/2018-annual-report.pdf?sfvrsn=6d769e1f_4
- Louisiana Department of Education. (2020). *Louisiana Early Childhood Quality and Progress: 2019-20 performance profiles*. https://www.louisianabelieves.com/docs/default-source/early-childhood/early-childhood-performance-profile-state-wide-webinar.pdf?sfvrsn=ae11911f_2
- Louisiana Early Childhood Education Act, S.B. 581, Legislature of Louisiana, Regular Session, 1 (2012). <http://www.legis.la.gov/legis/ViewDocument.aspx?d=800894&n=SB581%20Act%203>
- Louisiana Policy Institute for Children. (2014). *Louisiana Early Childhood Basics: Louisiana's Early Care and Education Programs*. https://0cd902dd-9de1-4dae-8781-4a355ebda8df.file.susur.com/ugd/43cca3_7beeed4b00654597bb5459de5cadedb0.pdf
- Markowitz, A. J., Bassok, D., & Hamre, B. (2018). Leveraging developmental insights to improve early childhood education. *Child Development Perspectives*, 12(2), 87–92. <https://doi.org/10.1111/cdep.12266>
- National Conference of State Legislatures. (2015). *2014 CCDBG Reauthorization Act: Implications and opportunities for state legislatures*. <https://www.ncsl.org/research/human-services/2014-ccdbg-reauthorization-act-implications-and-opportunities-for-state-legislatures.aspx>
- National Institute of Child Health and Human Development Early Child Care Research Network. (2002). Early child care and children's development prior to school entry: Results from the NICHD Study of Early Child Care. *American Educational Research Journal*, 39(1), 133–164. <https://doi.org/10.3102/00028312039001133>
- Office of Child Care. (2019a). *Child care and development fund statistics*. <https://www.acf.hhs.gov/occ/resource/ccdf-statistics>
- Office of Child Care. (2019b). *Preschool development grant birth through five grant competition*. U.S. Department of Health and Human Services & Administration for Children and Families. <https://www.acf.hhs.gov/occ/resource/pdg-b-5-initiative>
- Phillips, D. A., Lipsey, M. W., Dodge, K. A., Hasking, R., Bassok, D., Burchinal, M. R., Duncan, G. J., Dynarski, M., Magnuson, K. A., & Weiland, C. (2017, April 17). Puzzling it out: The current state of scientific knowledge on pre-kindergarten effects. *Brookings*. <https://www.brookings.edu/research/puzzling-it-out-the-current-state-of-scientific-knowledge-on-pre-kindergarten-effects/>
- 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(3), 281–303. [https://doi.org/10.1016/S0885-2006\(97\)90004-1](https://doi.org/10.1016/S0885-2006(97)90004-1)
- Pianta, R., & Howes, C. (Eds.). (2009). *The promise of pre-K*. Brookes.
- Pianta, R., La Paro, K. M., & Hamre, B. (2008). *Classroom Assessment Scoring System (CLASS) manual, pre-K*. Brookes.
- Regenstein, E., & Lipper, K. (2013). *A framework for choosing a state-level early childhood governance system*. BUILD Initiative. <http://www.buildinitiative.org/Portals/0/Uploads/Documents/Early%20Childhood%20Governance%20for%20Web.pdf>
- Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development*. National Academies Press. <http://www.nap.edu/catalog/9824>
- Tulane Institute of Early Childhood Mental Health & Louisiana Department of Health. (2016, Winter). *Early childhood risk and reach in Louisiana*. http://ldh.la.gov/assets/oph/Center-PHCH/Center-PH/maternal/RiskandReach_Report.pdf
- U.S. Department of Education. (2016, August). *2015 progress update: Race to the Top Early Learning Challenge*. <https://www2.ed.gov/programs/racetothetop-earlylearningchallenge/2015apr/2015progressfinal.pdf>
- U.S. Department of Health and Human Services. (2016). *Head Start federal funding and funded enrollment history*. <https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/head-start-federal-funding-funded-enrollment-history-eng.pdf>
- U.S. Department of Health and Human Services & Administration for Children and Families. (2010, January). *Head Start impact study: Final report*. https://www.acf.hhs.gov/sites/default/files/documents/opre/executive_summary_final_508.pdf
- Valentino, R. (2018). Will public pre-K really close achievement gaps? Gaps in prekindergarten quality between students and across states. *American Educational Research Journal*, 55(1), 79–116. <https://doi.org/10.3102/0002831217732000>
- Vitiello, V. E., Bassok, D., Hamre, B. K., Player, D., & Williford, A. P. (2018). Measuring the quality of teacher–child interactions at scale: Comparing research-based and state observation approaches. *Early Childhood Research Quarterly*, 44(3), 161–169. <https://doi.org/10.1016/j.ecresq.2018.03.003>
- Votruba-Drzal, E., Coley, R. L., Maldonado-Carreño, C., Li-Grining, C., & Chase-Lansdale, P. L. (2010). Child care and the development of behavior problems among economically disadvantaged children in middle childhood. *Child Development*, 81(5), 1460–1474. <https://doi.org/10.1111/j.1467-8624.2010.01485.x>
- Warner-Richter, M. (2016, February 9). *Promoting quality improvement in early care and education*. Child Trends. <https://www.childtrends.org/promoting-quality-improvement-in-early-care-and-education>
- Wasik, B. A., & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development. *Journal of Educational Psychology*, 103(2), 455–469. <https://doi.org/10.1037/a0023067>
- Wasik, B. A., & Hindman, A. H. (2020). Improving preschoolers' vocabulary development through a streamlined professional development intervention. *Early Childhood Research Quarterly*, 50(1), 101–113. <https://doi.org/10.1016/j.ecresq.2018.11.001>
- Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabu-

- lary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly*, 28(2), 199–209. <https://doi.org/10.1016/j.ecresq.2012.12.002>
- Whitebook, M., Howes, C., & Phillips, D. (1998). *Worthy work, unlivable wages: The National Child Care Staffing Study, 1988-1997*. Center for the Child Care Workforce. <http://eric.ed.gov/?id=ed419614>
- Whitebook, M., Phillips, D., & Howes, C. (2014). *Worthy work, STILL unlivable wages: The Early Childhood Workforce 25 years after the National Child Care Staffing Study*. Center for the Study of Child Care Employment, University of California, Berkeley. <https://cscce.berkeley.edu/wp-content/uploads/2014/ReportFINAL.pdf>
- Williford, A. P., Vick Whittaker, J. E., Vitiello, V. E., & Downer, J. T. (2013). Children’s engagement within the preschool classroom and their development of self-regulation. *Early Education & Development*, 24(2), 162–187.
- Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M., Espinosa, L. M., Gormley, W. T., Ludwig, J., Magnuson, K., Phillips, D., & Zaslow, M. J. (2013, October 11). *Investing in our future: The evidence base on preschool education*. Society for Research in Child Development: Foundation for Child Development. <https://www.fcd-us.org/the-evidence-base-on-preschool/>
- Zaslow, M., Anderson, R., Redd, Z., Wessel, J., Daneri, P., Green, K., Cavadel, E. W., Tarullo, L., Burchinal, M., & Martinez-Beck, I. (2016). Quality thresholds, features, and dosage in early care and education: Introduction and literature review. *Monographs of the Society for Research in Child Development*, 81(2), 7–26. <https://doi.org/10.1111/mono.12236>

Authors

DAPHNA BASSOK is an associate professor of education and public policy at the University of Virginia. She studies early childhood education policy.

PRESTON MAGOUIRK is a data scientist at Common App. He studies education policy.

ANNA J. MARKOWITZ is an assistant professor of human development and psychology at the School of Education and Information Sciences at the University of California, Los Angeles. Her work focuses on the ways in which policy affects the adults young children spend time with, and how this shapes children’s cognitive, social, and emotional development, particularly for children in families who are low-income or otherwise marginalized.