MEASURING UP

STRENGTHENING LOUISIANA'S EARLY CHILDHOOD ACCOUNTABILITY SYSTEM

APRIL 2016

LOUISIANA
POLICY INSTITUTE
for CHILDREN

MEASURING UP
Strengthening Louisiana's Early Childhood Accountability System

April 2016

Louisiana Policy Institute for Children
The Louisiana Policy Institute for Children is a non-partisan, non-profit, independent source of data, research and information for policy makers, stakeholders and the public at large around issues related to young children in Louisiana. The Institute develops policy proposals informed by research, best practices and the experiences of other states and performs educational and outreach activities around these recommended policy solutions.

You can find this report and additional information at www.policyinstitutela.org.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The Louisiana Early Childhood Education Act (Act 3) of 2012 requires the Board of Elementary and Secondary Education (BESE) and the Louisiana Department of Education (LDE) to establish a "uniform assessment and accountability system for publicly funded early childhood education programs." In the 2015-2016 school year, the Early Childhood Care and Education (ECCE) Accountability System established by the LDE and BESE in response to Act 3 will produce a practice rating of all of Louisiana's publicly funded ECCE programs, which serve 62,000 young children statewide, in one of three performance rating levels: Excellent, Proficient or Needs Improvement. These ratings will be based solely on one tool, the Classroom Assessment Scoring System (CLASS) observational assessment of classroom quality. CLASS is an assessment that measures emotional, behavioral, and instructional interactions among teachers and children; a trained observer observes and scores each classroom in four observational cycles, each lasting 20 minutes.

A great deal of research finds that CLASS scores are related to children's learning and development, meaning that the tool does an excellent job of measuring aspects of classroom quality that matter for children. Specifically, this research finds "domain-specific" associations of the Emotional Support and Classroom Organization Domains with children's positive social-emotional and behavioral outcomes, including children's social competence, behavioral regulation, and peer and teacher relationships. Similarly, the Instructional Support Domain has been found to be a critically important factor in children's learning and development across a wide range of cognitive domains and preacademic skills such as language, literacy, and math.

The 2015-2016 school year has been designated as a "Learning Year," in which the state has implemented CLASS observations and the ECCE Accountability System without imposing consequences for ECCE programs that demonstrate low quality. Importantly, the state is dedicated to using the findings from this Learning Year to adapt and improve future iterations of the system. Based on the results from 2015-2016, the LDE has proposed the following changes for the 2016-2017 school year, the first year in which the performance ratings will be associated with rewards and consequences:

1. Expand the ratings from three performance rating levels to four performance rating levels in order to better differentiate programs that are in the middle range. The four performance rating levels will be titled: Excellent, Proficient, Approaching Proficient, and Unsatisfactory.

2. Publicly recognize high-performing and quickly-improving programs through Honor Roll or Top Gains designations.

3. Prioritize the delivery of quality improvement supports to programs rated Unsatisfactory (less than 3.0 overall CLASS average scores).

4. Revoke public funding from programs rated Unsatisfactory for two or more years within a three-year period.

5. Maintain the use of third party observers' scores in lieu of local observers' scores when scores differ by more than one point within the same semester, and where consistent scoring discrepancies occur (defined as more than 50% of observer scores diverge by more than one point), use only third party scores for accountability purposes.

BESE's Early Care and Education Advisory Council will consider these proposed changes at its meeting on May 18, 2016, and BESE will consider them at its meetings on June 21 and 22, 2016.

The Louisiana Policy Institute for Children, with support from research expertise at the Ounce of Prevention Fund, has conducted a review of national research on CLASS and its use in accountability and Quality Rating and Improvement System (QRIS) across the country, examined data and findings from Louisiana's Learning Year, and scrutinized the LDE's proposed changes.

Based on our analyses, the Louisiana Policy Institute for Children supports the changes to the current system that the LDE has proposed in #1-5 above. These represent important steps towards ensuring that Louisiana's ECCE Accountability System is valid, reliable, and results in higher quality early childhood experiences for children statewide. However, additional modifications are necessary to ensure the integrity of the ECCE Accountability System and ensure that it drives ECCE program improvements not only among the lowest performing programs, but also among those that are performing just below a level of quality that would support school readiness for the children they serve. The authors of this report put forth the following recommendations for strengthening the ECCE Accountability System:

RECOMMENDATION 1: Adopt the LDE's proposal to introduce a fourth performance level, and revise the way the rating levels are computed to more meaningfully differentiate between levels of program quality, inform professional development, and incentivize quality improvement.

Challenge: As currently designed, programs are rated by averaging *all* of the domain scores of the CLASS assessment together to compute one score that determines the program's overall rating. This methodology of averaging all the domains together to create one score stands in contrast both to the academic research and the practices of every other state that uses the CLASS for rating purposes. Averaging the scores across all the domains allows a program's overall rating to mask an extremely low score in one domain. This is particularly problematic because recent research provides robust evidence that classrooms must score above a threshold of about 3.0 on the CLASS Instructional Support domain in order to successfully support children's preacademic learning.

Recommended Solutions:

- Calculate program Performance Ratings based on meeting minimum score thresholds in *both* of two CLASS score domain categories:

   1. Average of Emotional Support, Classroom Organization, and Emotional & Behavioral Supports Domains; and

   2. Average of Instructional Support and Engaged Supports for Learning Domains.

- Over the next three years, phase in research-based Proficiency thresholds at 5.0 for the Emotional Support/Classroom Organization/Emotional & Behavioral Supports domain category and at 3.0 Instructional Support/Engaged Supports for Learning.

- Given the inequities in funding and supports across the many ECCE programs held to the same standards by the ECCE Accountability System—with programs funded through the Child Care Assistance Program receiving substantially less funding and supports than either Head Start/Early Head Start or school systems—it will be critical to target supports such that the system drives improvement and does not penalize the most under-resourced programs.

RECOMMENDATION 2: Revise systems for CLASS observation to ensure that accountability scores are accurate, reliable, and efficiently obtained.

Challenge: The credibility of Louisiana's ECCE Accountability System hinges on the reliability of

observation scores. Currently, Louisiana requires each classroom to be observed twice per year by local observers—once in the Fall and once in the Spring. Many networks are using teachers, directors, or other employees of the schools or programs themselves to conduct the assessments (even of their own programs). These local observers are required to calibrate and establish inter-rater reliability with other observers within their own lead agency—but not with local observers in other lead agencies. That is, there is no system in place to ensure inter-rater reliability and calibration across networks statewide. Instead, the LDE has developed a system by which third party observers from the Picard Center at the University of Louisiana-Lafayette observe at least 50% of classrooms statewide one time per year. For each classroom, if the third party observer's domain scores differ from local observer's domain scores by more than one point in the same semester, then the third party's domain score replaces the local observer's domain score for accountability purposes. In Fall 2015, approximately one third (34%) of local observers' scores were replaced in this manner.

Recommended Solutions:

- Establish a system to ensure inter-rater reliability and calibration across all observers who contribute scores to the accountability system statewide—across all lead agencies in all geographic regions.

- Define reliable observers as those whose scores are within one point of third party scores at least 80% of the time. The LDE has proposed to define reliable local observers as those who score within one point of third party observations at least 50% of the time—well below the accepted standard of 80% of scores within one point established by the CLASS authors, and supported by research evidence.

- To further ensure inter-rater reliability among observers, deploy observers such that each classroom's accountability ratings are composed of scores from at least two different observers.

- Conduct at least one unannounced reliable observation per classroom per year for accountability purposes between November and March. Research suggests that CLASS scores are typically most stable in the middle of the school year – between November and March – and that they tend to decline rapidly in the last month of the school year.

- Do not conduct CLASS observations to be used for accountability purposes in the first or last month of the school year.

- To avoid penalizing programs for the month in which they were observed and to reward programs that have substantially improved their scores over the course of the year, calculate accountability scores based on each classroom's one highest score on each CLASS dimension (regardless of whether it was recorded in the Fall or Spring).

RECOMMENDATION 3: Strengthen and refine the system of targeted supports, incentives, and consequences that are aligned to the Performance Rating System to more effectively drive quality improvement.

Challenge: A rating system in and of itself will not drive improvement. Rather, to inspire improvement, it must operate within an ecosystem that uses the information provided by the rating system to make decisions. For parents, this would mean using Performance Ratings to help them choose among ECCE programs. For the state and school districts, this would mean using Performance Ratings to target supports and professional development to programs' specific strengths and weaknesses, especially for

those that receive disproportionately lower levels of funding (i.e. child care programs). Also for the state, this would include using Performance Ratings to deliver rewards to high-quality programs and consequences to eliminate persistently low-quality programs.

Recommended Solutions:

- Adopt the LDE's proposal to revoke public funding from programs rated Unsatisfactory for two or more years within a three-year period and redistribute these funds to the highest-quality programs to support their expansion.

- Conduct a needs analysis in any region where public funding is revoked from low-quality programs and ensure that children do not lose access to ECCE.

- Prioritize both Unsatisfactory *and* Approaching Proficient programs for intensive evidence-based mental health consultation services and targeted supports for curriculum implementation and instruction.

- Pair directors and teachers from Unsatisfactory and Approaching Proficient programs with directors and teachers from Excellent programs in a voluntary mentorship model.

- Over time, adopt the LDE's proposal to phase in a fifth performance level to inspire continuous improvement among all programs.

- Restructure the School Readiness Tax Credits to support and incentivize child care programs to improve classroom quality and improve their Performance Ratings under the new system.

RECOMMENDATION 4: Conduct an implementation evaluation to understand how the Rating and Improvement System is functioning and if it is achieving the goals outlined in Act 3.

Challenge: Conducting an implementation evaluation of the state's new system is a critical component to ensuring that the ECCE Accountability System is able to accomplish its goals. Specifically, this type of evaluation is needed to provide Louisiana with immediate feedback on the implementation of the new system in order to inform any course corrections needed to ensure that (1) Performance Ratings are valid and reliable, (2) the System meaningfully differentiates between levels of program quality, and (3) delivers a robust set of quality improvement supports and incentives for improvement, as well as consequences for failure to improve.

Recommended Solutions:

- Conduct an implementation evaluation of the state's new system to answer questions about whether the system:

    1. Is based on Performance Ratings that are valid and reliable;

    2. Meaningfully differentiates between levels of program quality; and

    3. Delivers a robust set of quality improvement supports and incentives for improvement, as well as consequences for failure to improve.

- Use findings from the implementation evaluation to inform the modification and improvement of the ECCE Accountability System in an iterative fashion.

RECOMMENDATION 5: Explore other key indicators of program quality that could be included in Performance Ratings in the future to more fully capture aspects of quality that support children's learning and development.

Challenge: Currently, Louisiana's ECCE Accountability System relies solely on CLASS to assign programs Performance Ratings. Although the LDE plans to provide additional information about program best practices on program Performance Profiles, they have proposed that these indicators not be included in the calculation of the program Performance Ratings that are the topic of our Recommendations 1-4. No other state in the country relies solely on one measure for the assignment of program Performance ratings in early childhood accountability or QRIS.

Recommended Solutions:

- Adopt the LDE's proposal to retain CLASS scores as the sole indicator used to calculate Performance Ratings in academic year 2016-17.

- Adopt the LDE's proposal to continue to include all other indicators of program quality on the ECCE Program Performance Profiles.

- Explore new and emerging tools—including the Five Essentials-Early Education and New York City's Pre-K for All Program Quality Standards Monitoring Tool— that measure other key aspects of program quality and consider their inclusion in calculating Performance Ratings in future years.

Conclusion

Since Act 3 passed in 2012, Louisiana has taken great strides to align standards and improve coordination and quality across its early childhood care and education programs in a historically disjointed and underfunded system. In order to accomplish this herculean task, one major lever that the state has chosen to employ is an ECCE Accountability System.

When developed and implemented in a thoughtful and evidence-based way, in tandem with sufficient resources and targeted supports, accountability systems can be a helpful way to hold all programs to the same standards, to communicate program performance, and to incentivize programs to continue improving. As Louisiana's ECCE Accountability System moves forward, it will be incumbent on program providers, parents, and stakeholders to ensure the reliability and integrity of the accountability system and to advocate to the state to use the results to drive change, including the strategic provision of targeted quality improvement supports and incentives, defunding consistently failing programs, and ensuring there are sufficient resources in the system to fund both the necessary infrastructure and the programs themselves.

This report provides a starting point for considering changes to Louisiana's ECCE Accountability System that build upon and strengthen the work that the LDE has already begun towards ensuring that all young children in Louisiana get a strong start in a high quality ECCE program.

BACKGROUND

Louisiana currently operates five types of publicly funded Early Childhood Care and Education (ECCE) programs. Four programs are administered by the state—the LA4 Prekindergarten Program, the 8(g) Student Enhancement Block Grant Program, the Nonpublic Schools Early Childhood Development (NSECD) Program, and the Child Care Assistance Program (CCAP). In addition to these programs, Louisiana also has Early Head Start/Head Start programs, which are federally funded and administered. These different programs are tied to specific funding streams linked to varying eligibility requirements for family participation, different licensing requirements, and variable measures for quality (see Table 1 for examples of variation). In the past, CCAP programs and Early Head Start/Head Start programs had the option to participate in the state's voluntary Quality Rating Improvement System (QRIS), but participation was not mandatory.

Table 1. Example of variable eligibility, licensing, and performance requirements for Louisiana's ECCE programs					
Program		**Eligibility**		**Child:Adult Ratios**	**Accountability System (prior to 2015)**
		Household Income	**Child Age**		
Public School	LA4	At or below 185% Federal Poverty Level (FPL)	4 years old	10:1	K-12 Accountability System
	8(g)		4 years old		
Nonpublic School	NSECD		4 years old		
Child Care	Child Care Assistance Program	At or below 55% of State Median Income	0-12 years old	5:1 (<1 year old) 7:1 (1 year old) 11:1 (2 years old) 13:1 (3 years old) 15:1 (4 years old	Voluntary QRIS
Head Start	Early Head Start	At or below 100% FPL (Note: A small % can be above FPL)	0-2 years old	4:1	
	Head Start	At or below 100% FPL (Note: A small % can be above FPL)	3-4 years old	17:2 (3 year-olds) 20:2 (4 year-olds)	

The variability and lack of coordination across ECCE programs in Louisiana led to a system that was incapable of accurately and efficiently identifying gaps and redundancies in service and which was difficult for parents and providers to navigate. In an attempt to highlight and address these challenges

with ECCE programs, in 2012 the Louisiana Legislature passed the Louisiana Early Childhood Education Act.

Act 3 of 2012 (Louisiana Early Childhood Education Act)

The Louisiana Early Childhood Education Act called for the Louisiana Department of Education (LDE), the State Board of Elementary and Secondary Education (BESE), the Department of Children and Family Services (DCFS), and the Department of Health and Hospitals (DHH) to unify and streamline the early childhood system to improve service delivery for Louisiana's families and to improve academic outcomes for Louisiana's children.

Specifically, the legislation required the LDE, in collaboration with other agencies, to create a system across the state with:

- A definition of Kindergarten readiness aligned to K-12 standards;

- Performance targets for children under the age of 3 and academic standards for Kindergarten readiness for 3- and 4-year-olds;

- Coordination of licensing regulations for child care facilities and requirements for participation in Louisiana's QRIS system; and

- A uniform assessment and accountability system for publicly-funded ECCE programs that includes letter grades indicative of student performance to be implemented by the 2015-2016 academic year.

As the LDE has begun work on the implementation of the Louisiana Early Childhood Education Act, the Louisiana state legislature has passed supplemental legislation in subsequent sessions that further define licensing regulations and coordinated enrollment systems. The legislature also transferred CCAP from DCFS to the LDE so that all ECCE programs could be housed under the same agency. In 2014, the state legislature also passed a resolution requiring BESE to develop a statewide model for ECCE funding and to report recommendations for funding ECCE to the legislature.

Implementation of Early Childhood Care and Education Networks

Louisiana's Early Childhood Education Act required the state to develop a statewide network to implement the changes set forth in the legislation. However, the legislation provided little direction around the structure of the network, other than to specify that it was to implement the regulations described above – Kindergarten readiness standards, performance targets, licensing regulation coordination, and an assessment and accountability system for ECCE. In order to accomplish these goals, the LDE decided to create locally based "community networks" to be led by lead agencies. In a few communities, these are the existing Resource and Referral Organizations. However, for most networks, the local school district is the lead agency. Community networks and lead agencies are written into BESE regulations, although "community network" is not formally defined or codified in Louisiana law. Further, no structure for local governance of the community network has been established in policy or law.

In 2013, the LDE spearheaded a pilot program of the initial 13 community networks and defined a vision that the community network pilots would work to ensure that all publicly funded ECCE programs had unified expectations, excellent teachers, equitable funding, access to quality options, and accountability and support (as described in Table 2).

Table 2. Unifying principles of Louisiana's ECCE community networks	
Principle	**Description**
Unified expectations	Shared high standards for what children should learn 0-5 and for what excellent teaching looks like
Excellent teachers	Teachers who are excellent at interacting with children and guiding learning
Equitable funding	Adequate funding levels across programs for programs to be able to serve children well
Access to quality options	A range of high-quality ECCE programs that families can easily evaluate and choose to apply to
Accountability and support	Targeted support based on a rating and reporting system that provides parents and the public with information about program quality

By 2014, the state launched 16 additional community networks, and by 2015 all communities across the state had lead agencies in place. As the networks launched, BESE worked to create and amend policies around the networks, including their required obligations around unified enrollment, evaluation, and accountability. All lead agencies are now working towards unified needs assessment and coordinated application and enrollment systems, as well as the accountability system, though the networks vary in their progress towards these goals.

LOUISIANA'S ECCE ACCOUNTABILITY SYSTEM

One of the lead agencies' primary tasks during the pilot phase has been to coordinate the evaluation of every program and every classroom in the network. Because the Louisiana Early Childhood Education Act required an accountability system for ECCE programs to be in place by the 2015-2016 academic year, BESE and the LDE developed a performance framework for evaluating ECCE programs and designated 2015-2016 a "Learning Year" in which there would be no consequences for programs that demonstrate low quality.

As part of its ECCE accountability system, the state created "Performance Profiles" intended to define core elements needed for positive child outcomes, provide simple and clear information about what was happening in classrooms, clearly articulate a path to improvement, and give families an easy way to compare choices in their community. Although Louisiana's ECCE Program Performance Profiles include information about classroom best practice—including the programs' child assessments, adult-to-child ratios, and teacher qualifications—BESE and the LDE chose to use the Classroom Assessment Scoring System (CLASS)[1] as the sole input for a program's overall quality rating across three levels—Needs Improvement, Proficient, and Excellent.

The Classroom Assessment Scoring System (CLASS)

The CLASS is an important measure of classroom quality in ECCE classrooms. This rigorously-tested observational measure has been used extensively in research and professional development systems for over a decade and has more recently become a top choice for use in QRIS throughout the country. In addition, the CLASS was also adopted by the federal Head Start program as part of its designation renewal system.

CLASS is an assessment tool that measures teacher-child interactions across various domains. In order to rate each classroom's CLASS performance through a formal observation, an observer must use the tool across four observation cycles (each lasting 20 minutes) to score each dimension on a 7-point scale. As illustrated in Table 3, the CLASS dimensions are then organized into domains—Emotional & Behavioral Support and Engaged Support for Learning in the CLASS-Toddler, and Emotional Support, Classroom Organization, and Instructional Support in the CLASS-PreK. Although the PreK CLASS tool measures three domains of teacher-child interactions, more recent and sophisticated measurement analysis on the CLASS-PreK tool suggests that in fact Emotional Support and Classroom Organization should be considered in tandem, while Instructional Support measures aspects of classroom quality that are quite distinct.[2] In particular, Instructional Support measures critical aspects of teaching such as the extent to which teachers engage in stimulating, language-rich interactions that help children acquire new knowledge and skills by providing challenging but accessible input to children and eliciting verbal responses and reactions from them.[3]

Table 3. Descriptions of domains and dimensions measured by CLASS		
Domain		**Dimension**
Toddler	Emotional and Behavioral Support	Positive Climate
		Negative Climate
		Teacher Sensitivity
		Regard for Child Perspectives
		Behavior Guidance
	Engaged Support for Learning	Facilitation of Learning and Development
		Quality of Feedback
		Language Modeling
Pre-K	Emotional Support	Positive Climate
		Negative Climate
		Teacher Sensitivity
		Regard for Student Perspectives
	Classroom Organization	Behavior Management
		Productivity
		Instructional Learning Formats
	Instructional Support	Concept Development
		Quality of Feedback
		Language Modeling

A great deal of research finds that CLASS scores are related to children's learning and development, meaning that the tool does a good job of measuring aspects of classroom quality that matter for children. Specifically, this research finds "domain-specific" associations of Emotional Support and Classroom Organization with children's positive social-emotional and behavioral outcomes, including children's social competence,[4] behavioral regulation,[5] and peer and teacher relationships.[6] In addition, Instructional Support has been found to be a critically important factor in children's learning and development across a wide range of cognitive domains and preacademic skills such as language, literacy, and math.[7] There is also some evidence that weaker "cross-domain" associations (e.g., the association of Emotional Support and Organizational Support with preacademic outcomes and vice versa) also exist,[8] suggesting that these three measures of quality may be particularly powerful in combination.[9] Because these domains are associated with different aspects of child development and learning, it is critical to ensure that programs achieve and maintain high quality on all CLASS domains.

Louisiana has chosen to task its lead agencies with the coordination of CLASS observations at the classroom and program levels. During the 2015-2016 academic year, the state required networks to observe each ECCE classroom twice: once in the Fall and once in the Spring. According to the LDE, most networks have been on track to successfully comply with these requirements, as 98% of publicly funded toddler and Prekindergarten classrooms were observed using the CLASS during Fall 2015.

In order to monitor and address issues of observer reliability, the LDE contracted with the Cecil J. Picard Center for Child Development and Lifelong Learning at the University of Louisiana-Lafayette to conduct one observation per year in at least 50% of all ECCE classrooms as third party observers. The LDE further regulated that in any classroom that was observed by both third party and a local observer in the same semester, the third party observer's scores would be used in lieu of the local observer's scores for any CLASS domain for which scores diverged by more than one point. Preliminary data from the LDE indicates that this domain level score replacement was 34% in Fall 2015.

In addition to being tasked with implementing a reliable CLASS observation system, the LDE also had to make decisions about how to calculate programs' Performance Ratings from resulting CLASS scores. Under the accountability system created by LDE, all CLASS dimension scores from Fall and Spring observations in all classrooms are averaged together to calculate an overall program Performance Rating based on one overall CLASS average score. For this first Learning Year, LDE has used a three level system, and will rate programs as Excellent, Proficient or Needs Assistance based on their overall CLASS scores, as described in Table 4.

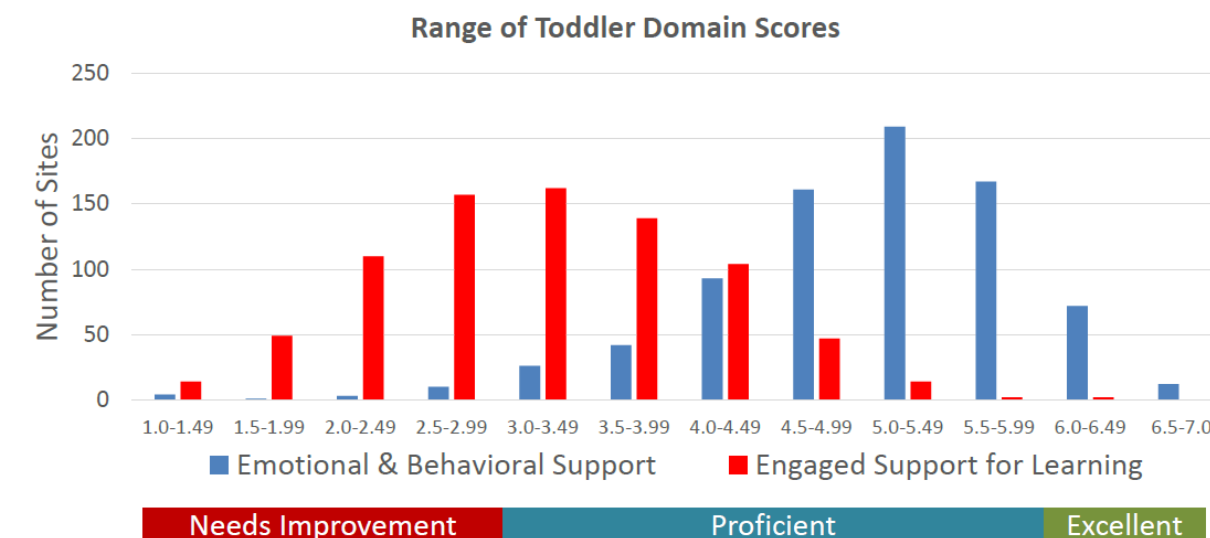| *Table 4.* CLASS Scores and Program Performance Ratings	
Performance Rating	**CLASS Score Range**
Excellent	≥6.00
Proficient	3.00-5.99
Needs Improvement	<3.00

Preliminary data from Fall 2015 CLASS observations indicate that an overwhelming proportion of classrooms (95.6%) were rated as Proficient using this system. Despite the large number of classrooms that scored in the Proficiency range, Fall 2015 data indicate that there is substantial variation across classrooms on individual domain scores, suggesting that within the Proficiency range, children are experiencing different levels of care and instruction (see Figures 1 and 2). In particular, it is notable that statewide the distribution of Instructional Support scores is substantially different than those of Emotional Support and Classroom Organization. That is, while virtually all programs in Louisiana are at least Proficient in the Emotional Support, Classroom Organization, and Emotional & Behavioral Supports domains, nearly half of programs are classified as Needs Improvement in Instructional Support or Engaged Supports for Learning. These data thus suggest that the current performance rating system—which classifies nearly all programs as Proficient—masks substantial variability in program quality.

Figure 1. Louisiana CLASS-PreK Scores in Fall 2015

Range of PreK Domain Scores

Number of Sites

Needs Improvement | Proficient | Excellent

■ Emotional Support   ■ Classroom Organization   ■ Instructional Support
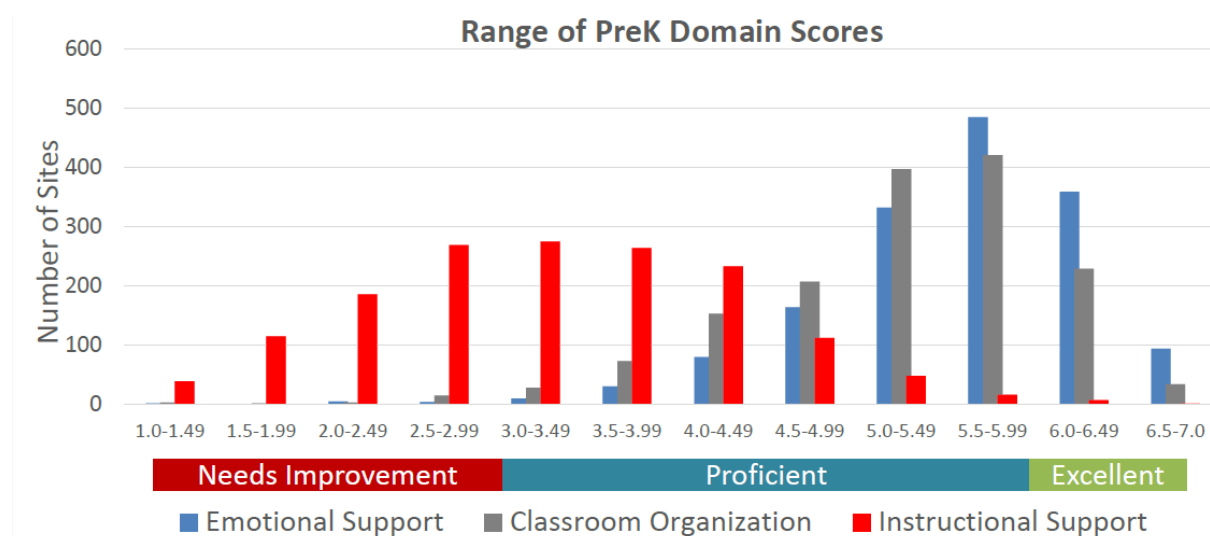
Source: Louisiana Department of Education, 2016

Figure 2. Louisiana CLASS-Toddler Scores in Fall 2015

Range of Toddler Domain Scores

Number of Sites

■ Emotional & Behavioral Support   ■ Engaged Support for Learning

Needs Improvement | Proficient | Excellent

Source: Louisiana Department of Education, 2016

LDE's Proposed Changes to the ECCE Accountability System

Importantly, the LDE has learned a great deal from the 2015-2016 Learning Year with the state's ECCE Accountability System. As the state moves forward with the ECCE Accountability System, the LDE has proposed the following changes for the 2016-2017 year, the first year for which the performance ratings will be associated with rewards and consequences:

- Expand the Performance Rating system from three performance levels to four performance levels in order to better differentiate among programs that fall in the middle range.

- Publicly recognize high-performing and quickly-improving programs through Honor Roll or Top Gains designations.

- Prioritize the delivery of quality improvement supports to programs rated Unsatisfactory (<3.0 overall CLASS average scores).

- Revoke public funding from programs rated Unsatisfactory for two or more years within a three-year period.

- Strengthen auditing of observer reliability, and maintain the use of third party scores in lieu of local scores when scores differ by more than one point within the same semester.

We applaud the LDE's thoughtful proposals and strongly support these proposed changes for the 2016-2017 year. In addition, as the state considers these changes to the ECCE Accountability System, we urge Louisiana to continue to keep research and precedent from other states in mind. The following section details recommendations that build upon the LDE's proposed changes and are based on best practices that have emerged from national CLASS research and research on QRIS practices nationwide.

RECOMMENDATIONS FOR IMPROVEMENTS TO LOUISIANA'S ECCE ACCOUNTABILITY SYSTEM

For Louisiana's ECCE Accountability System to be effective, it must (a) ensure the accuracy and comparability of CLASS scores and (b) meaningfully differentiate program quality. In order to ensure that the children of Louisiana benefit from high-quality early childhood experiences, the system must also (c) include a robust set of quality improvement supports and incentives for improvement, and consequences for failure to improve.

The changes to the current system that LDE has proposed for the 2016-17 academic year represent important steps towards ensuring that Louisiana's ECCE Accountability System is valid, reliable, and results in higher quality early childhood experiences for children statewide. However, additional modifications are necessary to ensure that the ECCE Accountability System meets these lofty goals and continues to drive ECCE program improvements.

Specifically, our recommendations are as follows. We discuss each in more detail below.

1. **Adopt the LDE's proposal to introduce a fourth performance level, and revise the way the rating levels are computed to more meaningfully differentiate between levels of program quality, inform professional development, and incentivize quality improvement.**

   o Calculate program Performance Ratings based on meeting minimum score thresholds in each of two CLASS score domain categories:

      1. Average of Emotional Support, Classroom Organization, and Emotional & Behavioral Supports dimensions; and

      2. Average of Instructional Support and Engaged Supports for Learning dimensions.

   o Over the next three years, phase in research-based Proficiency thresholds at 5.0 for the Emotional Support/Classroom Organization/Emotional & Behavioral Supports domain category and at 3.0 Instructional Support/Engaged Supports for Learning.

2. **Revise systems for CLASS observation to ensure that accountability scores are accurate, reliable, and efficiently obtained.**

   - Establish a system to ensure inter-rater reliability and calibration across all observers who contribute scores to the accountability system statewide—across all lead agencies in all geographic regions.

   - Define reliable observers as those whose scores are within one point of third party scores at least 80% of the time.

   - Deploy observers such that each classroom's accountability ratings are composed of scores from at least two different observers.

   - Calculate accountability scores based on each classroom's one highest score on each CLASS dimension (regardless of whether it was recorded in the Fall or Spring).

   - Conduct at least one unannounced reliable observation per classroom per year for accountability purposes between November and March.

   - Do not conduct CLASS observations to be used for accountability purposes in the first or last month of the school year.

3. **Strengthen and refine the system of targeted supports, incentives, and consequences that are aligned to the Performance Rating System to more effectively drive quality improvement.**

   - Adopt the LDE's proposal to revoke public funding from programs rated Unsatisfactory for two or more years within a three-year period and redistribute these funds to the highest-quality programs to support their expansion.

   - Conduct a needs analysis in any region where public funding is revoked from low-quality programs and ensure that children do not lose access to ECCE.

   - Prioritize both Unsatisfactory *and* Approaching Proficient programs for intensive evidence-based mental health consultation services and targeted supports for curriculum implementation and instruction.

   - Pair directors and teachers from Unsatisfactory and Approaching Proficient programs with directors and teachers from Excellent and Distinction programs in a voluntary mentorship model.

   - Over time, adopt the LDE's proposal to phase in a fifth performance level to inspire continuous improvement among all programs.

   - Restructure the School Readiness Tax Credits to support and incentivize child care programs to improve classroom quality and improve their Performance Ratings.

4. **Conduct an implementation evaluation to understand how the Rating and Improvement System is functioning and if it is achieving the goals outlined in Act 3.**

   - Conduct an implementation evaluation of the state's new system to answer questions about whether the system:

     1. Is based on Performance Ratings that are valid and reliable;

     2. Meaningfully differentiates between levels of program quality; and

3. Delivers a robust set of quality improvement supports and incentives for improvement, as well as consequences for failure to improve.

- Use findings from the implementation evaluation to inform the modification and improvement of the ECCE Accountability System in an iterative fashion.

5. **Explore other key indicators of program quality that could be included in Performance Ratings in the future to more fully capture aspects of quality that support children's learning and development.**

- Adopt the LDE's proposal to retain CLASS scores as the sole indicator used to calculate Performance Ratings in academic year 2016-17.

- Adopt the LDE's proposal to continue to include all other indicators of program quality on the ECCE Program Performance Profiles.

- Explore new and emerging tools—including the Five Essentials-Early Education and New York City's Pre-K for All Program Quality Standards Monitoring Tool— that measure other key aspects of program quality and consider their inclusion in calculating Performance Ratings in future years.

RECOMMENDATION 1: Adopt the LDE's proposal to introduce a fourth performance level, and revise the way the rating levels are computed to more meaningfully differentiate between levels of program quality, inform professional development, and incentivize quality improvement.

Currently, Louisiana assigns program performance ratings based on one overall CLASS average score, determined by averaging all of dimension scores from all of the classroom observations conducted within each program together to achieve a single score. In other words, a program's overall CLASS score is derived by averaging together all toddler and Pre-K observation results from both the Fall and Spring observation periods. In the 2015-2016 Learning Year, these overall CLASS scores were then classified into the Needs Improvement, Proficient, and Excellent Performance Ratings (see Table 4). The LDE has proposed two changes to this system in 2016-2017: (1) a shift from three to four performance levels, and (2) the elimination of Negative Climate dimension scores from calculations of overall CLASS average scores.

What does research say about CLASS scores and children's learning?

Since the CLASS tool was published, a great deal of research has explored thresholds in CLASS-PreK scores in an effort to better understand associations with children's learning and development. Overall, this research finds that a moderately high-quality classroom, as measured by the CLASS-PreK tool, is the minimum required to produce an association with positive preacademic, social-emotional, and behavioral outcomes for children.[10] Said another way, research suggests that preacademic outcomes—such as language, literacy, and math skills—are more strongly related to instructional support on CLASS when classrooms are in the moderate to high quality range than in the low quality range. Similarly, social-emotional and behavioral outcomes may be more strongly related to emotional support on CLASS when classrooms are in the high quality range rather than in the moderate or low quality range.

This research is clearest about instructional support, suggesting that there is no relationship between improvements in the quality of instructional support and improvements in child outcomes until quality reaches a certain level. *In other words, instructional support must reach a minimum threshold before it*

has any impact on children's learning. Once quality crosses that "threshold," child outcomes improve as quality improves. Although this research does not decisively point one precise threshold, recent studies do provide robust evidence to suggest that it falls in the range of 2.75 to 3.25[11]—in line with Louisiana's current definition of a Proficient Performance Rating at 3.0.

The research is not quite as clear on whether such thresholds exist for Emotional Support and Classroom Organization. While some studies find that a similar threshold pattern exists for these dimensions as well, others find that improvements in these dimensions are related to improvement in child outcomes across the entire range of scores—even in classrooms within the moderate to low quality range. That is, improvements in emotional support and classroom organizational may always by related to improvements in child outcomes. In general, research has placed the cut-point for acceptably-high quality in these domains between 5.0 and 5.5—much higher than Louisiana's current definition of a "proficient" level of quality, which is 3.0.

How does Louisiana's method for calculating CLASS scores for program performance ratings differ from all other states across the country?

Of the 39 state QRIS profiled in the *2014 Catalog and Comparison of Quality Rating and Improvement Systems* (referred to as the *"Compendium"*),[12] 19 incorporate use of the CLASS. Of those, 13 QRIS incorporate CLASS scores either by establishing a minimum average score (or range of scores) separately for each CLASS domain that programs must meet or exceed in order to be awarded specified rating levels or points. In the six state QRIS that reported using more than one CLASS instrument (Pre-K, Infant and/or Toddler), these QRIS use the same minimum score for each relevant domain across the three instruments. The domain-level scores are then averaged across age groups to determine one average score across all age groups for each domain.

Unlike Louisiana, none of the 13 QRIS combine the domain-level averages into an overall average CLASS score across all three domains to determine the program ratings. Instead, all 13 specify separate minimum score thresholds for each domain. For example, rather than defining the rating of "Proficient" as equal to an overall CLASS score of 4.5 by averaging together all dimension scores across the domains of Emotional Support, Classroom Organization and Instructional Support, "Proficient" would instead be defined as scoring at or above a 5.0 on Emotional Support, *and* a 5.0 on Classroom Organization, *and* a 3.0 on Instructional Support (see Appendix A for more details).

What does this mean for Louisiana?

To review, research and Louisiana's own data show that:

1. The Emotional Support and Classroom Organization domains of the CLASS-PreK tool measure similar aspects of quality, but Instructional Support measures very different kinds of teaching practices;

2. The threshold at which domain scores are most strongly related to child outcomes is around 3.0 on Instructional Support versus around 5.0 on Emotional Support and Classroom Organization. We can assume that thresholds on the CLASS-Toddler tool are likely similar;

3. Nationwide, all 13 QRIS that include CLASS scores set separate minimum score thresholds or ranges for each CLASS domain;

4. Programs tend to score lower on Instructional Support and Engaged Supports for Learning than they do on other domains, which means that overall CLASS average scores can mask variation in domain-level scores.

This leads us to conclude that performance ratings cannot meaningfully distinguish between levels of program quality when they are derived from one overall average score. *Louisiana's current method of averaging dimension scores across all CLASS domains to calculate an overall CLASS average score and program Performance Rating stands in contrast both to the academic research and common practices in other states.*

In order to address this issue, we recommend that Louisiana revise the ECCE Accountability System to adhere to the following criteria:

1. Calculate Performance Rating based on two separate CLASS score domain categories for each program:

   a. Average of Emotional Support, Classroom Organization, and Emotional & Behavioral Supports dimensions;

   b. Average of Instructional Support and Engaged Supports for Learning dimensions;

2. Set Minimum Score thresholds for the Emotional Support/Classroom Organization/Emotional & Behavioral Supports domain category that are distinct from those set for Instructional Support/Engaged Supports for Learning;

3. Require programs to score above the Minimum Scores on *both* domain categories (Emotional Support/Classroom Organization/Emotional & Behavioral Supports *and* Instructional Support/Engaged Supports for Learning) in order to be rated at each Performance Level; and

4. Over the next three years, phase in Minimum Score thresholds that are scaled based on research evidence such that the "Proficient" Performance Level is equal to the level of quality necessary to support children's learning.

The LDE has proposed to establish a four-level Performance Rating System in academic year 2016-17, which we strongly support, and in addition, we recommend that the state move to a five-level system in the future. Below we recommend target Minimum Score thresholds for four- and five-level systems that both adhere to the four criteria laid about above. *In order to allow time for programs to learn more about the CLASS and work to improve their classroom quality, we recommend that the target Minimum Score thresholds detailed in the four-level system be phased in over a period of three years.* That is, we recommend that these target Minimum Score thresholds be fully implemented by the 2019-2020 year— the same year by which lead teachers in all publicly-funded ECCE programs in Louisiana will be required to hold a credential.[13] Further, we recommend that the five-level system be introduced only once the recommended target Minimum Score thresholds have been fully implemented in the four-level system.

Specifically, based on the research reviewed above, both recommended target systems reflect our understanding that *it is critical for programs to achieve at least at 3.00 in the Instructional Support/Engaged Supports for Learning domain category to be considered Proficient.* Further, because research shows that children's learning grows as scores in the Instructional Support domain increase, the recommended systems incentivize continued improvement on Instructional Support/Engaged Support for Learning even above a 3.0. In addition, given that Louisiana programs already tend to score

higher in the Emotional Support and Classroom Organization domains, the recommended target systems are designed such that *programs must achieve at least at 5.0 on Emotional Support/Classroom Organization/Emotional & Behavioral Supports domain category to be considered Proficient*.

Four Performance Levels

The recommended target four-level system sets the bar for "Proficient" at the thresholds at which research shows that classroom practices begin to support children's learning. However, above those minimum thresholds, research shows that even higher levels of quality benefit children even more. At "Excellent," the target system sets the bar high—creating incentives for programs to continue to improve beyond that minimum threshold. In order to be rated at each Performance Level, a program would be required to achieve average CLASS scores at or above the Minimum Score threshold on *both* domain categories (see Table 5 and Figure 3). *To reiterate, we recommend that these target Minimum Score thresholds be phased in over a period of three years, such that they are fully implemented by the 2019-2020 year.*

Table 5. Proposed Four Performance Level System				
	Unsatisfactory	**Approaching Proficient**	**Proficient**	**Excellent**
Emotional Support/Classroom Organization/Emotional & Behavioral Supports	<4.00	4.00-4.99	5.00-5.99	≥6.00
Instructional Support/Engaged Support for Learning	<2.00	2.00-2.99	3.00-3.99	≥4.00

Based on preliminary data from Fall 2015 provided by the LDE, we estimate that under this target four-level rating scale and Louisiana's current system for combining local and third party data collection, the average Fall 2015 program in Louisiana would be rated at Proficient in both domain categories on both the CLASS-PreK and CLASS-Toddler tools, for an overall performance rating of Proficient. Based on these same data, we estimate that—using only CLASS-PreK scores—the average school-based and Early Head Start programs would also be rated overall at Proficient, but the average child care and Head Start programs would be rated at Approaching Proficient overall (with the Head Start average right on the cusp of Proficient). Finally, we estimate that—using only CLASS-Toddler scores—the average Early Head Start, Head Start, and child care infant-toddler programs would be rated overall at Proficient, but the average school-based programs would be rated at Approaching Proficient. For context, nationwide, the average Head Start program would be rated Approaching Proficient.[14] The average New York City and Boston public Pre-K programs would be rated Proficient, although average Instructional Support scores in Boston fall within the Excellent range (4.30).[15]

Figure 3. Proposed Four Performance Level System							
	Average Emotional Support/Classroom Organization/Emotional & Behavioral Support						
		1	2	3	4	5	6	7

Figure 3. Proposed Four Performance Level System (matrix of Average Instructional Support/Engaged Support for Learning (rows 1–7) against Average Emotional Support/Classroom Organization/Emotional & Behavioral Support (columns 1–7)).

Key

	Unsatisfactory
	Approaching Proficient
	Proficient
	Excellent

Five Performance Levels

Because the Early Childhood Education Act requires the ECCE Accountability System to ultimately use letter grades as Performance Ratings, eventually the state may choose to move to a five-level rating scale (and we recommend that they do). One way to think about the addition of a Performance Rating level is to "raise the bar" by adding an additional level ("Distinction") above the Excellent level in the target four-level system (see Table 6 and Figure 4). Once programs have become acquainted with the four-level scale and have been given ample opportunity to improve their rating within it, the addition of a higher top-tier rating could provide an important incentive for high-quality programs to continue to improve. *To reiterate, we recommend that the five-level system be introduced only once the recommended target Minimum Score thresholds have been fully implemented in the four-level system.*

Table 6. Proposed Five Performance Level System					
	Unsatisfactory	Approaching Proficient	Proficient	Excellent	Distinction
Emotional Support/Classroom Organization/Emotional & Behavioral Supports	<4.00	4.00-4.99	5.00-5.99	6.00-6.49	≥6.50
Instructional Support/Engaged Support for Learning	<2.00	2.00-2.99	3.00-3.99	4.00-4.49	≥4.50

Figure 4. Proposed Five Performance Level System

Key
- Unsatisfactory
- Approaching Proficient
- Proficient
- Excellent
- Distinction

As in the four-level scale, this target system sets the bar for Proficient around the thresholds where research shows that classroom practices begin to be related to children's learning. At Distinction, it raises the bar even higher—creating additional incentives for even the highest-quality programs to continue to improve.

RECOMMENDATION 2: Revise systems for CLASS observation to ensure that accountability scores are accurate, reliable, and efficiently obtained.

Currently, Louisiana requires each classroom to be observed twice per year—once in the Fall and once in the Spring. Lead agencies are tasked with organizing the observation process and are required to use CLASS-certified observers. Many lead agencies are using teachers, directors, or other employees of the schools or programs themselves to conduct the assessments (even of their own program). Some networks have also chosen to recruit, train, and pay outside observers who are not employed by the ECCE programs. Additionally, as described above in the background section, the LDE has also contracted with the Picard Center at the University of Louisiana-Lafayette to serve as third party observers. The Picard Center is required to observe at least 50% of classrooms in Louisiana once per year.

The LDE has given lead agencies the task of rater calibration, allocating funds to them and requiring them to develop procedures related to shadow scoring, calibration, inter-rater reliability, and conflict of interest. Each lead agency is required to conduct their own inter-rater reliability observation checks in at least 10 percent of all classrooms observed, but there is no requirement for a minimum number of checks for each observer.

In addition, the LDE has also developed a system by which third party scores take precedence over local scores. In the current system, in any classroom that was observed by both third party and a local observer in the same semester, the third party observer's scores are used in lieu of the local observer's scores for any CLASS domain for which scores that diverged by more than one point. The use of third party scores across classrooms and networks ensures that there is at least some calibration across classrooms and networks, and that conflicts of interest are diminished.

Louisiana faces substantial challenges with observer reliability.

Still, the LDE has indicated that some local observers continue to face challenges with reliability. In fact, LDE's own analyses reveal that approximately one third (34%) of local observers' scores differ from third party observers' domain scores by more than one point—the established margin of error for calculating inter-rater reliability—within the same semester. Thus, the state has proposed that if a local observer's Fall scores are found to differ from third party scores by more than one point consistently—more than 50% of the time—across multiple domains in multiple observations, then the observer will not be allowed to observe in the Spring. Additionally, if several observations are found to differ by more than one point on a repeated basis, then a program's Performance Rating would be assigned based completely on third party observers' scores.

The credibility of Louisiana's ECCE Accountability System hinges on the reliability of observation scores. While the LDE's important adjustments begin to address issues of observer reliability, we posit that additional, more stringent measures are needed. First, the LDE's proposed system relies on the assumption that the third party observers' scores are always more accurate than the local observers'. This assumption should be tested and verified to ensure that replacing local observers' scores (rather than, for example, averaging them together) does in fact increase score accuracy. Second, even if it is

determined that LDE's score replacement system is justified, the current system is not set up to allow for the replacement of all of the unreliable scores produced by local observers. Local observers conduct two observations in each classroom per year (one in the Fall and one in the Spring), while third party observers only conduct one audit per year in at least 50% of classrooms (half of them in the Fall and the other half in the Spring). This means that at a minimum, third party observers audit only 25% of all local observations—25% of Fall observations and 25% of Spring observations. In Fall 2015, approximately one third (34%) of the domain scores resulting from these audited observations differed by more than one point and were thus replaced by third party observers' scores. This means that while 34% of audited scores were replaced, at a minimum, only 8.5% (that is, 34% of 25%) of *all* local observers' Fall domain scores were replaced. If we assuming that classrooms audited in the fall were chosen at random and that replacement rates are found to be similar among Spring domain scores, the same logic would apply to Spring observations—we would expect that 8.5% of *all* local observers' Spring domain scores will be replaced as well.

Therefore, this rough illustrative example suggests that if audited, 34% of *all* local observers' domain scores would differ from third party scores by more than one point and thus should be replaced. However, under the current system as few as a quarter of those—8.5% of *all* local observers' domain scores—may actually be replaced.

Importantly, this projection is likely an *underestimate* of the unreliable scores that will actually be replaced in 2015-2016 for two reasons. First, the LDE has indicated that over the course of the 2015-2016 school year, third party observers are in fact auditing 70%—rather than 50%—of classrooms. Adjusting our projection based on this 70% audit rate would increase our expectation of the proportion of unreliable scores currently being replaced from one quarter to roughly one third. Second, as described in more detail below, the LDE has several mechanisms in place to use reliability information from Fall audits to improve the reliability of Spring observations. To the extent that these mechanisms prove successful, our assumption that replacement rates among Spring domain scores will be similar to those found in the Fall is likely incorrect. However, even if we further adjust our projection to also include the generous assumption that these mechanisms are able to cut Spring replacement rates in half (from 34% to 17%), we would still expect the current system to fail to replace nearly one third of local observers' unreliable domain scores.

How can Louisiana do to address these observer reliability challenges?

When the CLASS tool is used for accountability purposes, it is critical that the data collection system be implemented in a way that is highly attentive to inter-rater reliability. As such, it is critical that the data collection system is designed to be highly attentive to ongoing reliability concerns. *This means that the system must be highly systematic about the ways in which it (1) trains, (2) calibrates, and (3) deploys observers to score classrooms.*

As described above, the LDE currently has systems in place to ensure that observers are trained to CLASS reliability standards (i.e. passes a reliability test in which they code videos within one point of the established Teachstone master codes). *We agree with the LDE that these standards should be maintained as they currently exist.* In addition, as described above, the LDE currently has standards in place to ensure that local observers maintain inter-reliability with *some* of their peers who contribute scores to the state's accountability system—with other local observers within their own lead agency, and with third party observers. The LDE has proposed to define reliable local observers as those who score within one point of third party observations at least 50% of the time.

There are two issues with Louisiana's current observer reliability standards and protocols. First, the 50% requirement for reliable observers is quite low—well below the accepted standard of 80% of scores within one point established by Teachstone, and supported by research evidence.[16] *As such, we recommend that Louisiana increase its 50% inter-rater reliability requirement to 80%.* That is, local observers would be required to score within one point of third party observer scores at least 80% of the time in order to be considered reliable and allowed to continue to contribute scores to the ECCE Accountability System. Such a change is an important step in increasing the reliability of the system overall. However, it does nothing to address another characteristic of Louisiana's current observer reliability system that also compromises inter-rater reliability. Currently, observers are required to calibrate and establish inter-rater reliability with other observers within their own lead agency—but not with local observers in other lead agencies. That is, there is no system in place to ensure inter-rater reliability and calibration across networks statewide. Such a system is critical in order to ensure that accountability ratings given based on scores from one lead agency are reliably equivalent to those given based on scores from another lead agency. *We recommend that Louisiana establish a system that requires and supports local observers to calibrate scores and establish inter-rater reliability with one another statewide—across all lead agencies in all geographic regions.*

However, even once a robust inter-rater reliability system is fully implemented, an observer's use of the CLASS tool is somewhat subjective—scores need only to be within one point to be considered reliable. Because an accountability system is built on the assumption that scores are measured in a comparable way across all classrooms, it is critical that unintended—but inevitable—observer bias be distributed as evenly as possible across all classrooms. One key to doing so is to *deploy observers such that each classroom's accountability ratings are composed of scores from at least two different observers.* Currently, the LDE deploys local observers to independently observe each classroom. As described above, third party observers then independently observe a sub-sample of classrooms at each site at a separate point in time. Under this system, most classrooms' ratings are based entirely on scores from one observer. However, alternate methods of deploying observers to classrooms could remedy both of these issues and move Louisiana closer to an accountability system that minimizes the impact of observer bias and maximizes rating reliability system-wide. Recommended options include:

1. Sending a team of at least two observers to simultaneously observe each program. Each observer starts in a different classroom and partway through the observation (e.g., after two cycles), observers switch classrooms. Because each observation lasts four cycles, each classroom's final score for the day would thus be a result of observations from multiple observers. This means that the impact of each observer's unintended bias on the classroom's final rating is minimized because it is balanced out by other observer's scores. However, we recognize that sending multiple observers to each program may be too costly to be feasible in Louisiana—particularly in small programs and rural areas.

2. Using video recordings to score classrooms. In this less expensive alternative, programs record and submit observation-length (i.e., 2 hours) video recordings of each classroom—recorded in compliance with state guidelines and standards (e.g., recorded on a tripod from a corner of the classroom in one take—no editing—within one week of a "surprise" notification). A team of observers would then watch and score these videos as they would in-person at the site. As in Option 1, observers would switch videos partway through such that each video is scored by at least two observers. As described above, this method would minimize the impact of each observer's unintended bias on the classroom's final rating. Because this option does not involve

the amount of travel required by Option 1, it is likely far less costly. This option could be implemented in a variety of ways, including:

    a. Centrally—Programs would submit videos to one central state agency to be scored by a team of observers. This team of observers may include local as well as third party observers. In this scenario, videos would be assigned to observers without regard for geographic location.

    b. Regionally—Programs would submit videos to their regional lead agency to be scored by a local team of observers. Each team may include local as well as third party observers. In this scenario, videos would be assigned to observers based on geographic location.

With either option it is critical that information regarding individual classrooms' strengths and weaknesses be shared back with the teacher and program to inform quality improvement efforts and professional development. Although it is important that the person sharing that information be well-versed in the CLASS tool (i.e., trained to reliability), it is *not* necessary for that person to be the same one who collected it. That is, the observer whose scores were used for accountability purposes does not need to be the one to communicate that information back to the program. In fact, local program staff who have been trained to reliability on the CLASS tool may be particularly well-positioned to share observation results, given that they likely have ongoing relationships with the program staff and are familiar with the local context, history, opportunities, and constraints.

When should CLASS observations be conducted?

Unlike Louisiana, which requires one Fall and one Spring observation per year in all classrooms, most QRIS across the country observe only a portion of the classrooms in each age group (e.g., one third of preK classrooms, one third of toddler classrooms) once throughout the year. Research suggests that CLASS scores are typically most stable in the middle of the year—between November and March[17]—and that they tend to decline rapidly in the last month of the school year.[18] This means that some variation in scores is likely attributed to when within the fall and spring months a classroom was observed. Louisiana's current ECCE Accountability System averages scores across observations, potentially penalizing a classroom for being observed in September rather than November or May rather than March. In addition, averaging scores across observations provides a good estimate of the average level of quality present in that classroom across the entire school year, but minimizes the potential rewards for a program that has substantially improved their scores over the course of the year. As such, if Louisiana chooses to continue to conduct two observations per classroom per year for accountability purposes, *we recommend that the classroom's one highest score on each CLASS dimension (regardless of whether it was recorded in the Fall or Spring) be used to calculate the accountability score.*

Furthermore, in order to capture classroom quality when it is most stable, *we recommend that Louisiana conduct at least one unannounced reliable observation (as described above) per classroom per year for accountability purposes between November and March*. In addition, *we recommend that no CLASS observations to be used for accountability purposes be conducted in the first or last month of the school year, when CLASS scores tend to be lower.*

One of the most critical attributes of an ECCE Accountability System is its ability to drive quality improvement. A rating system in and of itself will not drive the improvement. Rather, to inspire improvement, it must operate within an ecosystem that uses the information provided by the rating system to make decisions. For parents, this would mean using Performance Ratings to help them choose among ECCE programs. For networks, this would mean using Performance Ratings to target supports and professional development to programs' specific strengths and weaknesses. For the state, this would include using Performance Ratings to deliver rewards to high-quality programs and consequences to eliminate persistently low-quality programs.

Supports and Consequences for Low-Quality Programs

Any ECCE Accountability System must include appropriate consequences and incentives to help low quality programs move out of Unsatisfactory status and to encourage high quality programs to achieve and maintain Excellent or Distinction status. The LDE has proposed to terminate public funding for all programs that receive Unsatisfactory ratings two or more years within a three-year period, beginning in 2019. During that time, the state proposes to support programs to improve by providing 6 months of intensive evidence-based mental health consultation services to Unsatisfactory programs. In order to ensure that publicly-funded programs in Louisiana are not providing unsatisfactory levels of quality to young children, w*e strongly support the adoption of both of these proposals.*

However, we posit that these proposed supports and consequences may not be sufficient to ensure that Louisiana's young children experience high levels of quality statewide. In addition to the LDE's proposals, *we recommend that programs with an Approaching Proficient Performance Rating also receive priority status for mental health consultation services.* In particular, *we recommend that mental health consultation services be targeted specifically to those programs who need them most—those that score particularly low on Negative Climate and/or the Emotional Support/Classroom Organization/Emotional & Behavioral Supports domain category*. Finally, in order to provide greater levels of support to Approaching Proficient programs, which may be more ready and able to benefit from quality improvement supports than lower quality programs, we recommend that they be paired with high-quality programs with Excellent or Distinction ratings so that Approaching Proficient programs can learn from experts.

Importantly, evidence suggests that the most promising strategy for raising classroom quality—particularly the quality of Instructional Support—is intensive, relationship-based professional development aligned to an evidence-based, developmentally focused curriculum(a). In such models, an expert teacher provides feedback and support for in-classroom practice, either in person or in some cases through observation of videos of classroom teaching.[19] In order to support programs to improve the quality of the Instructional Support that they provide to young children, *we recommend that both Unsatisfactory and Approaching Proficient programs receive priority status for evidence-based, targeted supports for curriculum implementation and instruction.* In particular, *we recommend that curriculum implementation and instruction supports be targeted specifically to those programs who need them most—those that score particularly low on the Instructional Support/Engaged Support for Learning domain category*.

In addition, *we recommend that Louisiana restructure the School Readiness Tax Credits to support and incentivize child care programs to improve classroom quality and improve their Performance Ratings under the new system.*

Finally, we recommend that a needs analysis be conducted in any region where public funding is revoked from low-quality programs. Furthermore, we recommend that children enrolled in programs forced to close receive priority access to other publicly funded programs in the area to ensure that they do not lose access to ECCE altogether.

Rewarding High-Quality Programs

Above a minimum threshold, children's learning increases as classroom quality improves across the full range of the CLASS tool. As such, it is important to incentivize improvement not only at the bottom end of the distribution, but also in higher quality programs. The LDE has proposed that Excellent programs receive statewide recognition through publicly published Honor Rolls. *We wholeheartedly support the adoption of this proposal.* In addition, in order to support the expansion of the state's highest-quality ECCE programs, *we recommend that public funds revoked from Unsatisfactory programs be reallocated to Excellent programs (and subsequently to Distinction programs) that are willing and able to expand the number of slots they can offer to children*. This reallocation would help to incentivize high quality programs to achieve and maintain Excellent or Distinction status and would also help to (1) increase the percentage of publicly-funded slots in high quality programs and (2) mitigate any loss of system-wide capacity that often results from the closure of low-quality programs.[20] Finally, to further support program improvement *we recommend that directors and teachers from Unsatisfactory and Approaching Proficient programs be paired with directors and teachers from Excellent and Distinction programs in a voluntary mentorship model.*

RECOMMENDATION 4: Conduct an implementation evaluation to understand how the Rating and Improvement System is functioning and if it is achieving the goals outlined in Act 3.

As articulated by the LDE, the goals of Louisiana's Rating and Improvement System are to:

1. Define and measure core elements needed for positive child outcomes;
2. Provide simple, clear information about what's happening in classrooms;
3. Clearly articulate a path to improvement; and
4. Give families an easy way to compare choices in their community.[21]

Conducting an implementation evaluation of the state's new system is a critical component to ensuring that it is able to accomplish these goals. Specifically, this type of evaluation would provide Louisiana with immediate feedback on the implementation of the new system in order to inform any course corrections needed to ensure that (1) Performance Ratings are valid and reliable, (2) the System meaningfully differentiates between levels of program quality, and (3) delivers a robust set of quality improvement supports and incentives for improvement, as well as consequences for failure to improve.

Under the current proposed System for 2016-17, this means first focusing such an evaluation primarily on the implementation of the CLASS. *We recommend that Louisiana invest in an implementation evaluation that would answer questions such as*:

- Does the System ensure that the CLASS is administered by trained and reliable observers?

- Does the System ensure that reliability is assessed frequently through double-coded observations that are compared and discussed to minimize "drift" across the state?

- Are Minimum Score thresholds defined such that programs in each Performance Level are meaningfully similar to other programs rated at the same Level and meaningfully different from programs rated at other Levels?

Should the evaluation reveal the answer to any of these questions is no—or finds that feasible changes would improve the ability of the system to meet the stated goals—Louisiana should use that information to modify and improve its ECCE Accountability System accordingly.

Once Louisiana determines that the system is being implemented as intended, an important aspect of validating this new system is to assess the extent to which the ratings meaningfully differentiate higher-quality programs from lower-quality programs. If a system goal is to support children's development and learning, then it may also be useful to determine the extent to which children's development and learning is greater in higher-rated programs compared to lower-rated programs. Other questions of interest may include:

- To what extent are programs improving quality?

- Are families choosing care settings based on the ratings?

- Do families have the option to choose different care settings based on the performance ratings?

- Have program/provider morale and motivation to improve changed over time?

RECOMMENDATION 5: Explore other key indicators of program quality that could be included in Performance Ratings in the future to more fully capture aspects of quality that support children's learning and development.

Currently, Louisiana's ECCE Accountability System relies solely on CLASS to assign programs Performance Ratings. Although the LDE plans to provide additional information about program best practices on program Performance Profiles, they have proposed that these indicators not be included in the calculation of the program Performance Ratings that are the topic of our Recommendations 1-4.

No other state in the country relies solely on one measure for the assignment of program Performance ratings in QRIS. The 2014 *Compendium*,[22] profiles 39 QRIS efforts nationwide to measure, monitor, and promote high-quality child care. State QRIS administrators completed a survey in which they described several aspects of their QRIS, including information on indicators of quality, funding, incentives, technical assistance, observation tools, and others. Table 7 reports the number of states that included at least one indicator related to the identified quality components in their QRIS.

Table 7. Number and percentage of QRIS (n=39) incorporating specified quality components for center-based care (2014).

Administration and management	38 (97%)
Staff qualifications	
• Director qualifications	38 (97%)
• Teacher qualifications	37 (95%)
Family partnerships	37 (95%)
Environment	36 (92%)
Health and safety	34 (87%)
Professional Development	
• Director professional development	34 (87%)
• Teacher professional development	29 (74%)
Accreditation	33 (85%)
Curriculum	32 (82%)
Licensing compliance	32 (82%)
Ratio and group size	30 (77%)
Provisions for children with special needs	27 (69%)
Continuous quality improvement practices	22 (56%)
Child assessment	21 (54%)
Cultural/linguistic diversity	18 (46%)
Community involvement	17 (44%)
Interactions	12 (31%)

Source: *Catalog and Comparison of Quality Rating and Improvement Systems (2014). Available at:* http://qriscompendium.org/

Indeed, maintaining high levels of quality in a classrooms is not easy: even the best teachers require a great deal of training, experience, resources, and ongoing professional development and support to do so consistently and successfully.[23] Other states have built these elements into their QRIS. The following sections provide brief research summaries of key components of high-quality early care and education settings included in other states' QRIS that research show to be most strongly associated with high-quality teacher-child interactions and positive outcomes for children.

Administration and Management

Program management and administration (which includes features such as program leadership, compensation and benefits, program planning and evaluation, technology, human resources development, and fiscal management) can influence overall program quality in important ways,[24] particularly as it may influence the recruitment and retention of highly-qualified teachers/caregivers.[25] Research in K-12 education demonstrates that children who attend schools with low levels of teacher turnover demonstrate higher academic performance than children who attend schools with high levels of teacher turnover.[26] Programs can increase their chances of recruiting and retaining qualified staff by providing a professional work environment,[27] providing adequate compensation and paid planning time[28] and engaging teachers/caregivers in opportunities for performance evaluation and professional development to continuously improve their practice.[29]

Staff Qualifications

The education and training of individual teachers and caregivers have been consistent predictors of positive outcomes for children in multiple research studies.[30] However, some studies have not detected relationships (or have found contradictory relationships) between educational qualifications such as college degrees, classroom quality and children's outcomes.[31] The field widely acknowledges that most existing studies do not capture important features of pre-service education that may lead to mixed results, such as content of an early childhood degree, participation in field experiences, and capacity of degree-granting institutions.[32] Although the evidence is mixed, some studies suggest that a bachelor's degree with a focus on core competencies has the potential to offer early care and education providers the tools for establishing high-quality classrooms and promoting positive child outcomes, when other important supportive factors for providers are also in place.[33]

Professional Development

Structured professional development opportunities that focus on both knowledge and practice are often most effective in supporting teacher and caregivers practices.[34] This includes coupling on-site technical assistance with professional development efforts, such as a coach, so that early care and education providers are able to practice and implement what they have learned through trainings or workshops.[35] Professional development that is intensive, on-going, and includes targeted coaching can contribute to positive outcomes for children.[36]

Curriculum

An evidence-based, developmentally focused curriculum that supports the knowledge, skills, and abilities across all of the domains of early childhood development can help young children achieve goals that are educationally significant.[37] When an intensive developmentally focused curriculum is implemented to fidelity and in a way that supports children's conceptual understanding, children are most likely to achieve the intended goals and outcomes of that curriculum approach.[38] Training and supporting teachers on the use of specific curricula and providing onsite coaching has been found to increase the likelihood that curricula are implemented to fidelity.[39] The use of ongoing authentic observational assessment practices to plan and guide instruction with young children is widely acknowledged as a key component of a high-quality early care and education program.[40] When curriculum and assessment practices are aligned, developmentally appropriate, and implemented with fidelity the two practices can work together to support the growth and development of individual children and their needs.[41]

Promising New Tools to Measure These Elements

One key difference between these QRIS and Louisiana's ECCE Accountability System is that other states' QRIS are voluntary and do not carry the accountability consequences (i.e., defunding) that will be implemented in Louisiana. When such high-stakes consequences are tied to a Performance Rating, the level of rigor that must be applied in determining that rating is substantially higher than would be necessary in lower-stakes systems. Unfortunately, no rigorous measures of these important components of quality currently exist for use in an accountability system like the one that Louisiana is creating.

Therefore, *we recommend that the LDE's proposal to rely on CLASS scores as the sole indicator used to calculate Performance Ratings be adopted for the 2016-2017 year.* In addition, *we support the adoption*

of the LDE's proposal to report other key indicators of program quality on ECCE Program Performance Profiles. Furthermore, *we recommend that these Profiles be formatted to emphasize the importance of multiple measures of key components of high-quality early care and education settings that research show to be most strongly associated with high-quality teacher-child interactions and positive outcomes for children* (see above).

However, two tools that would be appropriate in this context—the Five Essentials-Early Education and New York City's Pre-K for All Program Quality Standards Monitoring Tool—are currently under development. Each of these tools represents an important complement to CLASS scores, and would broaden the definition of quality currently implicit in Louisiana's ECCE Accountability System and Performance Ratings.

Five Essentials-Early Education[42]
Research in elementary schools led to the identification of five school-level organizational conditions that are critical in supporting quality improvement. These conditions, as articulated in the Five Essentials Framework, are: Effective Leadership, Collaborative Teachers, Family Involvement, Supportive Environment, and Ambitious Instruction. [43] Each of the five essentials supports is uniquely important to a school's success, but they also work together to support professional relationships, classroom interactions, teaching, and learning. In fact, elementary schools strong in three or more of these supports were 10 times more likely than schools weak in most supports to substantially improve students' math and reading outcomes and attendance. Furthermore, a weakness in just one of these supports over time undermined other change efforts, reducing the probability of improvement to 10% or less.[44] Although widespread focus on these organizational conditions is relatively new compared to the field's focus on quality inside the classroom, both New York City and Illinois have been monitoring and holding schools accountable for the quality of these elements in all public schools for over a decade using valid and reliable surveys that are universally-distributed to teachers, (in New York City) parents, and (in later grades) students each year.[45] A new version of the tool (teacher and parent surveys) designed specifically for use in Pre-K programs is currently being developed and validated.

New York City Pre-K for All Program Quality Standards Monitoring Tool
The New York City Department of Education (NYCDOE) had developed a set of quality standards that define a vision for high-quality Pre-K for All programs in NYC. These standards describe the key practices of family engagement, rigorous and developmentally appropriate instruction, professional collaborations, and leadership that support children in gaining the knowledge and skills. They are aligned with the New York State Prekindergarten Foundation for the Common Core and with The Framework for Great Schools, which is similar to the Five Essentials and guides school improvement efforts across the pre-K to 12 continuum. [46] Currently, NYCDOE is working to develop a companion "snapshot" rubric that NYCDOE staff can use to measure key program quality standards in Pre-K for All programs in both public schools and community-based settings. This new tool will be used to make accountability decisions and to inform technical assistance efforts (including decisions about whether programs could benefit from the addition of an evidence-based curriculum and how much coaching might be needed).
We recommend that Louisiana continue to explore new and emerging tools that measure other key aspects of ECCE quality with the goal of incorporating valid and reliable measures of aspects such as organizational conditions and curriculum implementation into the Performance Rating in future years. Specifically, we recommend that LDE cultivate relationships with the developers of the Five Essentials-

Early Education and New York City's Pre-K for All Program Quality Standards Monitoring Tool to track their development closely and fully assess the potential of their application in Louisiana.

CONCLUSION

Since Act 3 passed in 2012, Louisiana has taken great strides to align standards and improve coordination and quality across its early childhood care and education programs in a historically disjointed and underfunded system. In order to accomplish this herculean task, one major lever that the state has chosen to employ is an ECCE Accountability System.

When developed and implemented in a thoughtful and evidence-based way, in tandem with sufficient resources and targeted supports, accountability systems can be a helpful way to hold all programs to the same standards, to communicate program performance, and to incentivize programs to continue improving. As Louisiana's ECCE Accountability System moves forward, it will be incumbent on program providers, parents, and stakeholders to ensure the reliability and integrity of the accountability system and to advocate to the state to use the results to drive change, including the strategic provision of targeted quality improvement supports and incentives and defunding consistently failing programs. It will also be critical that those programs, such as those funded through the Child Care Assistance Program, that currently receive the fewest resources receive sufficient supports to be successful in a system where they are held to the same standards as programs funded at higher levels.

This report provides a starting point for considering changes to Louisiana's ECCE Accountability System by building upon and strengthening the work that the LDE has already begun. With these recommendations and the input of ECCE stakeholders across the state, Louisiana can ensure that all young children in Louisiana get a strong start in a high-quality ECCE program.

Use of CLASS in QRIS across the country was identified by the 2014 *Catalog and Comparison of Quality Rating and Improvement Systems (Compendium).*[47] As a result, please note that the number of states using CLASS and the associated thresholds may have changed since the release of the Compendium.

Arizona

STAR Level	Emotional Support	Classroom Organization	Instructional Support
3	4.5	4.5	2.0
4	5.0	5.0	2.5
5	6.0	6.0	3.0

Colorado

Points Awarded	Emotional Support	Classroom Organization	Instructional Support
2	4.5	4.5	2.0
4	5.5	5.0	2.5
6	6.0	6.0	3.0

Florida – Guiding Stars of Duval

n/a	Emotional Support	Classroom Organization	Instructional Support
	4.0-4.49	3.5-3.99	2.0-2.15
	4.5-4.99	4-4.49	2.16-2.3
	5.0-5.49	4.5-4.99	2.31-2.75
	5.5-5.99	5-5.49	2.76-3.25
	>6.0	>5.5	>3.26

Florida – Palm Beach

Level	Emotional Support	Classroom Organization	Instructional Support
1*	<4.5	<3.5	<1.0
2*	4.5-4.99	3.5-3.99	1.0-1.49
3*	5.0	4.0	1.5
4*	6.5	5.5	3.0

For Level 1 and 2 programs only need to obtain the specified threshold for one of the subscales. For Levels 3 and 4 programs need to obtain the specified thresholds on all three subscales.

Florida – Miami Dade

Points Awarded	Emotional Support	Classroom Organization	Instructional Support
2	4.5	4.5	2.0
4	5.5	5.0	2.5
6	6.0	6.0	3.0

Illinois

Level	Emotional Support	Classroom Organization	Instructional Support
Silver	4.0-5.0 with no classroom below 3.0	4.0-5.0 with no classroom below 3.0	≥2.5
Gold	5.0 with no classroom below 4.0	5.0 with no classroom below 4.0	≥3.0

Minnesota

Points Awarded	Emotional Support	Classroom Organization	Instructional Support
0pts	1-3.99	1-2.59	1-2.49
.5ots	4-5.59	2.6-3.59	2.5-2.99
1 pt	5.6-7	3.6-7.0	3.0-7.0

Nebraska

Points Awarded	Emotional Support	Classroom Organization	Instructional Support
1pt	4.0-4.9	4.0-4.9	2.0-3.4
2pt	5.0-5.9	5.0-5.9	3.5-4.9
3pt	≥6.0	≥6.0	≥5.0

North Dakota (coming soon)

Level	Emotional Support	Classroom Organization	Instructional Support
Step 3	4	3	2

Oregon

Level	Emotional Support	Classroom Organization	Instructional Support
5 Star	5	5	5

Virginia

Level	Emotional Support	Classroom Organization	Instructional Support
4	5	5	3.25
5	6	6	4.25

Vermont

Points Awarded	Emotional Support	Classroom Organization	Instructional Support
4pts	≥4	≥4	≥4

Washington

Points Awarded	Emotional Support	Classroom Organization	Instructional Support
10pt	3.5-4.9	3.5-4.9	2-3.4
15pt	5.0-5.9	5.0-5.9	3.5-4.4
20pt	≥6.0	≥6.0	≥4.5

REFERENCES

[1] Pianta, R. C., LaParo, K. M., & Hamre, B. K. (2008). *Classroom assessment scoring system [CLASS] manual: Pre-K.* Baltimore, MD: Paul H. Brookes.

[2] Hamre, B., Hatfield, B., Pianta, R., & Jamil, F. (2014). Evidence for General and Domain-Specific Elements of Teacher–Child Interactions: Associations With Preschool Children's Development. Child development, 85(3), 1257-1274.

[3] Mashburn, A.J., Pianta, R.C., Hamre, B.K., Downer, J.T., Barbarin, O., Bryant, D., … Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*, 732–749.; Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M.R., Espinosa, L.M., Gormley, W.T., … Zaslow, M.J. (2013). *Investing in our future: The evidence base on preschool education.* Policy brief, Society for Research in Child Development and Foundation for Child Development. Retrieved from: http://fcd-us.org/resources/evidence-base-preschool

[4] Curby, T. W., LoCasale-Crouch, J., Konold, T. R., Pianta, R. C., Howes, C., Burchinal, M., & Barbarin, O. (2009). The relations of observed pre-K classroom quality profiles to children's achievement and social competence. *Early Education and Development, 20*(2), 346–372.

Mashburn, A. J., Pianta, R. C., Hamre, B. K., Downer, J. T., Barbarin, O., Bryant, D., &Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79*, 732–749.

[5] Mashburn et al., 2008.

[6] Hamre, B. K., & Pianta, R. C. (2005). Can instructional and emotional support in the first grade classroom make a difference for children at risk of school failure? *Child Development, 76*, 949–967.

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

[7] Curby et al., 2009; Mashburn et al., 2008; Howes, C., Burchinal, M., Pianta, R., Bryant, D., Early, D., Clifford, R., & Barbarin, O.(2008). Ready to learn? Children's pre-academic achievement in pre-kindergarten programs. *Early Childhood Research Quarterly, 23*, 27–50.

[8] Pianta et al., 2002; Pianta, R. C., Belsky, J., Vandergrift, N., Houts, R., & Morrison, F. J. (2008). Classroom effects on children's achievement trajectories in elementary school. *American Educational Research Journal, 45*, 365–397.

[9] Downer, J., Sabol, T. J., & Hamre, B. (2010). Teacher–child interactions in the classroom: toward a theory of within-and cross-domain links to children's developmental outcomes. *Early Education and Development, 21*(5), 699–723.

[10] Burchinal, M., Vernon-Feagans, L., Vitiello, V., Greenberg, M., & Family Life Project Key Investigators. (2014). Thresholds in the association between child care quality and child outcomes in rural preschool children. *Early childhood research quarterly*, *29*(1), 41-51.

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.

Burchinal, M., Zaslow, M., & Tarullo, L. (Unpublished manuscript) *Quality Thresholds, Features, and Dosage in Early Care and Education: Secondary Data Analyses of Child Outcomes.*

Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly*, *28*(2), 199-209.

[11] Burchinal et al., unpublished manuscript.

[12] The Build Initiative & Child Trends. (2014). A Catalog and Comparison of Quality Rating and Improvement Systems (QRIS) [Data System]. Retrieved from http://qriscompendium.org/ on May 26, 2015.

[13] Louisiana Department of Education. (2015). *Early Childhood Teaching Policies.* Retrieved from: https://www.louisianabelieves.com/docs/default-source/early-childhood/ecce-ac---12-8-presentation-on-proposed-teacher-policies.pdf?sfvrsn=2

[14] Office of Head Start. (2015). *A National Overview of Grantee CLASS Scores in 2015.* Retrieved from: http://eclkc.ohs.acf.hhs.gov/hslc/data/class-reports/class-data-2015.html

Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013).

[15] New York City Department of Education. (2015). *Pre-K Program Assessments Classroom Assessment Scoring System (CLASS) and Early Childhood Environmental Rating Scale – Revised (ECERS-R) Release.* Retrieved from: http://schools.nyc.gov/NR/rdonlyres/A8A27BFE-7C58-4F03-8EB7-B90E01BA3D0D/0/CLASSandECERSRReleaseDeckFinal.pdf

[16] Landis, J. R., & Kotch, G. G. (1977). The measurement of observer agreement forcategorical data. Biometrics, 33, 159–174.

[17] Pianta, R., Mashburn, A., Downer, J., Hamre, B., & Justice, L. (2008). Effects of web-mediated professional development resources on teacher–child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 23*(4), 431–451.

[18] Pianta, R. C., & Hamre, B. K. (2009). Conceptualization, measurement, and improvement of classroom processes: Standardized observation can leverage capacity. *Educational researcher*, *38*(2), 109-119.

[19] Yoshikawa et al. (2013).

[20] Hotz, V. J., & Xiao, M. (2011). The impact of regulations on the supply and quality of care in child care markets. *The American economic review*, *101*(5), 1775.

[21] Louisiana Department of Education. (2016). *Early Childhood Roundtables: What to Expect in 2016-17.* Retrieved from: http://www.louisianabelieves.com/docs/default-source/early-childhood/2016-early-childhood-roundtable-presentation.pdf?sfvrsn=2

[22] The Build Initiative & Child Trends. (2014). A Catalog and Comparison of Quality Rating and Improvement Systems (QRIS) [Data System]. Retrieved from http://qriscompendium.org/ on May 26, 2015.

[23] Early, D., Maxwell, K., Burchinal, M., Alva, S., Bender, R., Bryant, D., … & 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.

[24] Bloom, P. J., & Sheerer, M. (1992). The effect of leadership training on child care program quality. *Early Childhood Research Quarterly*, *7*(4), 579–594. http://doi.org/10.1016/0885-2006(92)90112-C

McCormick Center for Early Childhood Leadership. (2010). *Head Start administrative practices, director qualifications, and links to program quality.* Wheeling, Illinois: National-Louis University.

McCormick Center for Early Childhood Leadership. (2012). *Leadership Matters*. Wheeling, Illinois: National-Louis University. Retrieved from http://mccormickcenter.nl.edu/leadership-matters-report/

[25] Torquati, J. C., Raikes, H., & Huddleston-Casas, C. a. (2007). Teacher education, motivation, compensation, workplace support, and links to quality of center-based child care and teachers' intention to stay in the early childhood profession. *Early Childhood Research Quarterly*, *22*(2), 261–275. http://doi.org/10.1016/j.ecresq.2007.03.004

[26] Grissmer, D., & Kirby, S. (1997). Teacher turnover and teacher quality. *The Teachers College Record*, *99*(1), 45-56.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American educational research journal*, *38*(3), 499-534.

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal*, *50*(1), 4-36.

[27] Ryan, S., & Whitebook, M. (2012). *More than teachers: The early care and education workforce*. In R. C. Pianta (Ed.), Handbook of early childhood education (pp. 92-110). New York: The Guilford Press.

[28] Sosinsky, L. S., & Gilliam, W. S. (2011). Assistant teachers in prekindergarten programs: What roles do lead teachers feel assistants play in classroom management and teaching? *Early Education and Development*, *22*(4), 676–706. http://doi.org/10.1080/10409289.2010.497432

National Association for the Education of Young Children & National Association of Child Care Resource and Referral Agencies. (2011). *Early Childhood Education Professional Development: Training and Technical Assistance Glossary.* Washington, DC. Retrieved from http://www.naeyc.org/GlossaryTraining_TA.pdf

[29] National Association for the Education of Young Children & National Association of Early Childhood Specialists in State Departments of Education. (2003). Early Childhood Curriculum, Assessment, and Program Evaluation: Building an Effective, Accountable System in Programs for Children Birth through Age 8. *Young Children*, (November). Retrieved from https://www.naeyc.org/files/naeyc/file/positions/pscape.pdf

[30] Institute Of Medicine & National Research Council. (2012). *The Early Childhood Care and Education Workforce: Challenges and Opportunities: A Workshop Report*. Washington, DC: The National Academies Press.

Institute of Medicine (IOM) and National Research Council (NRC). 2015. *Transforming the workforce for children birth through age 8: A unifying foundation.* Washington, DC: The National Academies Press.

National Institute of Child Health and Human Development Early Child Care Research Network, & Duncan, G. (2003a). Does quality of child care affect child outcomes at age 4 ½? *Developmental Psychology, 39*(3), 451-469.

National Institute of Child Health and Human Development Early Child Care Research Network & Duncan, G. (2003b). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development, 74*(5), 1454–1475.

Tout, K., Zaslow, M., & Berry, D. (2006). *Quality and qualifications: Links between professional development and quality in early care and education settings.* In Martha Zaslow & Ivelisse Martinez-Beck (Eds.), Critical issues in early childhood professional development (pp. 77-110). Baltimore, MD: Paul H. Brooks.

Whitebook, M. & Ryan, S. (2011) *Degrees in context: Asking the right questions about preparing skilled and effective teachers of young children*. Preschool Policy Brief, 22. Berkeley, CA: Center for the Study of Child Care Employment and the National Institute for Early Education Research.

[31] Early, D. M., Maxwell, K. L., Burchinal, M., Bender, R. H., Ebanks, C., Henry, G. T., Iriondo-Perez,

J., Mashburn, A. J., Pianta, R. C., Alva, S., Bryant, D., Cai, K., Clifford, R. M., Griffin, J. A., Howes, C., Jeon, H., Peisner-Feinberg, E., 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.

[32] Whitebook, M. & Ryan, S. (2011) *Degrees in context: Asking the right questions about preparing skilled and effective teachers of young children*. Preschool Policy Brief, 22. Berkeley, CA: Center for the Study of Child Care Employment and the National Institute for Early Education Research.

[33] Institute of Medicine (IOM) and National Research Council (NRC). 2015. *Transforming the workforce for children birth through age 8: A unifying foundation.* Washington, DC: The National Academies Press.

[34] Tout, K., Halle, T., Zaslow, M., & Starr, R. (2009). *Evaluation of the Early Childhood Educator*

Professional Development Program: Final Report. Washington, DC: Report prepared for the U.S. Department of Education.

[35] Tout, K., Epstein, D., Soli, M., & Lowe, C. (2015). A Blueprint for Early Care and Education

Quality Improvement Initiatives: Final Report. Bethesda, M.D.: Child Trends.

[36] Hindman, A. H., Erhart, A. C., & Wasik, B. A. (2012). Reducing the matthew effect: Lessons from the ExCELL head start intervention. *Early Education and Development, 23*(5), 781-806. doi:http://dx.doi.org/10.1080/10409289.2010.549443

National Professional Development Center on Inclusion. (2010). *Building integrated professional development systems in early childhood: Recommendations for states.* Chapel Hill: The University of North Carolina, FPG Child Development Institute.

Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T.,

Ludwig, J., Magnuson, K. A., ... & Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education (Vol. 9)*. Society for Research in Child Development and Foundation for Child Development.

[37] National Association for the Education of Young Children. (2009). *Developmentally Appropriate Practice in Early Childhood Programs Serving Children from Birth through Age 8.* Washington, DC. Retrieved from http://www.naeyc.org/files/naeyc/file/positions/PSDAP.pdf

[38] Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W. T.,

Ludwig, J., Magnuson, K. A., ... & Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education (Vol. 9)*. Society for Research in Child Development and Foundation for Child Development.

[39] Nikki Aikens and Lauren Akers, Background Review of Existing Literature on Coaching (Washington, DC: Mathematica Policy Research, 2011).

[40] National Research Council. (2008). Early childhood assessment: Why, what, and how. Committee on Developmental Outcomes and assessments for Young Children, C. E. Snow and S. B. Van Hemel, Eds. Board on Children, Youth, and Families, Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

[41] Chen, J. Q., & McNamee, G. (2006). Strengthening early childhood teacher preparation: Integrating assessment, curriculum development, and instructional practice in student teaching. *Journal of Early Childhood Teacher Education*, *27*(2), 109-128.

Schilder, D., & Carolan, M. (2014). State of the states policy snapshot: State early childhood assessment policies.

[42] Pacchiano, D., Ehrlich, S., Stein, A., & Luppescu, S. (2015). Five Essentials—Early Education Measurement System. Chicago, IL: Ounce of Prevention Fund & University of Chicago Consortium on Chicago School Research.

[43] Bryk, A. S., Sebring, P. B., Allensworth, E., Easton, J. Q., & Luppescu, S. (2010). *Organizing schools for improvement: Lessons from Chicago*. University of Chicago Press.

[44] Ibid.

[45] Nathanson, L., McCormick, M., and Kemple, J.J. (2013). *Strengthening Assessments of School Climate: Lessons From the NYC School Survey*. New York: Research Alliance for New York City Schools; UChicago Impact. (2016). *Background, Predictive Validity, and Reliability: 5Essentials*. UChicago Impact. http://uchicagoimpact.org/5essentials/background.

[46] New York City Department of Education. (2016). Pre-K for All Program Quality Standards. Retrieved from: http://schools.nyc.gov/Academics/EarlyChildhood/support/PKAPQS.htm

[47] The Build Initiative & Child Trends. (2014). A Catalog and Comparison of Quality Rating and Improvement Systems (QRIS) [Data System]. Retrieved from http://qriscompendium.org/ on May 26, 2015.