Report to the Mississippi Legislature

2019 Report on the Early Learning Collaborative Act of 2013: An Evaluation of the Operation and Effectiveness of the Program

#640
December 16, 2019

PEER MISSISSIPPI
Joint Legislative Committee on Performance Evaluation and Expenditure Review

PEER: The Mississippi Legislature's Oversight Agency

The Mississippi Legislature created the Joint Legislative Committee on Performance Evaluation and Expenditure Review (PEER Committee) by statute in 1973. A joint committee, the PEER Committee is composed of seven members of the House of Representatives appointed by the Speaker and seven members of the Senate appointed by the Lieutenant Governor. Appointments are made for four-year terms, with one Senator and one Representative appointed from each of the U.S. Congressional Districts and three at-large members appointed from each house. Committee officers are elected by the membership, with officers alternating annually between the two houses. All Committee actions by statute require a majority vote of four Representatives and four Senators voting in the affirmative.

Mississippi's constitution gives the Legislature broad power to conduct examinations and investigations. PEER is authorized by law to review any public entity, including contractors supported in whole or in part by public funds, and to address any issues that may require legislative action. PEER has statutory access to all state and local records and has subpoena power to compel testimony or the production of documents.

PEER provides a variety of services to the Legislature, including program evaluations, economy and efficiency reviews, financial audits, limited scope evaluations, fiscal notes, special investigations, briefings to individual legislators, testimony, and other governmental research and assistance. The Committee identifies inefficiency or ineffectiveness or a failure to accomplish legislative objectives, and makes recommendations for redefinition, redirection, redistribution and/or restructuring of Mississippi government. As directed by and subject to the prior approval of the PEER Committee, the Committee's professional staff executes audit and evaluation projects obtaining information and developing options for consideration by the Committee. The PEER Committee releases reports to the Legislature, Governor, Lieutenant Governor, and the agency examined.

The Committee assigns top priority to written requests from individual legislators and legislative committees. The Committee also considers PEER staff proposals and written requests from state officials and others.

PEER Committee
Post Office Box 1204
Jackson, MS 39215-1204

(Tel.) 601-359-1226
(Fax) 601-359-1420
(Website) www.peer.ms.gov

The Mississippi Legislature

Joint Committee on Performance Evaluation and Expenditure Review

PEER Committee

SENATORS
LYDIA CHASSANIOL
Vice Chair
KEVIN BLACKWELL
Secretary
TERRY C. BURTON
GARY JACKSON
SAMPSON JACKSON II
CHARLES YOUNGER

REPRESENTATIVES
BECKY CURRIE
Chair
RICHARD BENNETT
STEVE HORNE
TIMMY LADNER
MARGARET ELLIS ROGERS
RAY ROGERS
PERCY W. WATSON

TELEPHONE:
(601) 359-1226

FAX:
(601) 359-1420

Post Office Box 1204
Jackson, Mississippi 39215-1204

James A. Barber
Executive Director

www.peer.ms.gov

OFFICES:
Woolfolk Building, Suite 301-A
501 North West Street
Jackson, Mississippi 39201

January 6, 2020

Honorable Tate Reeves, Governor
Honorable Delbert Hosemann, Lieutenant Governor
Honorable Philip Gunn, Speaker of the House
Members of the Mississippi State Legislature

On December 16, 2019, the PEER Committee authorized release of the report titled ***2019 Report on the Early Learning Collaborative Act of 2013: An Evaluation of the Operation and Effectiveness of the Program.***

Representative Becky Currie, Chair

This report does not recommend increased funding or additional staff.

Table of Contents

List of Exhibits

PEER MISSISSIPPI
Joint Legislative Committee on Performance Evaluation and Expenditure Review

Joint Legislative Committee on Performance Evaluation and Expenditure Review
Report Highlights
December 16, 2019

2019 Report on the Early Learning Collaborative Act of 2013: An Evaluation of the Operations and Effectiveness of the Program

CONCLUSION: Since program implementation in FY 2014, the Mississippi Department of Education (MDE) has improved its ability to measure prekindergarten program effectiveness by adding two valid tests to its evaluation. However, MDE's method for evaluating effectiveness fails to adequately measure collaborative or site success. Also, MDE requires collaboratives to use a curriculum that does not have research evidence to support its effectiveness in improving student learning. PEER's independent evaluation of scores on each of MDE's tests more accurately measures program and site effectiveness and provides valuable diagnostic information for MDE and the collaboratives. For example, PEER's evaluation of Brigance III scores found that 25 sites (39%) improved from the pre-test to the post-test in the 2018-19 school year by a statistically significant amount (i.e., to a degree distinguishable from chance). Also, PEER's evaluation of Classroom Assessment Scoring System (CLASS) scores for the 2018-19 school year found that collaborative classrooms performed extremely well on the Emotional Support domain when compared nationally, performed close to the national norms on the Classroom Organization domain, and performed worse than the national norms on the Instructional Support domain.

Purpose and Scope:

In its FY 2021 budget request, the Mississippi Department of Education (MDE) is seeking an additional $3.2 million for the state's prekindergarten program, bringing the total program funding to $9.9 million annually. Thus, the Legislature will need information for the 2020 Regular Session to determine whether to expand the program by appropriating the requested additional funds.

This is the second evaluation of the prekindergarten program conducted by PEER, with the first evaluation conducted in 2015.

Background:

In its 2013 Regular Session, the Legislature enacted the "Early Learning Collaborative Act of 2013," which directs MDE to implement a voluntary prekindergarten program in the state on a phased-in basis. The prekindergarten program's purpose is to help ensure that all children have access to quality early childhood education and development services.

Mississippi's prekindergarten program received legislative appropriations of $14.5 million from FY 2017-19. This funding assisted fourteen early learning collaboratives (including 62 sites) in implementing prekindergarten programs that served approximately 2,220 students in the 2018-19 school year.

Based on MDE's evaluation of collaboratives and sites for the 2018-19 school year, MDE rated 59 sites successful and placed three sites on probation.

Research evidence suggests that public prekindergarten programs have a positive short-term impact on children's cognitive abilities and their readiness for kindergarten. However, there is not sufficient evidence to conclude that public prekindergarten programs have a positive longer-term impact. Thus, it is essential for the state to develop a research strategy to demonstrate prekindergarten program impacts, and it is incumbent upon MDE to identify the best approaches to use in early learning collaborative classrooms that will lead to positive short and long-term impacts.

Improved Ability to Measure Program Success

In school year 2014-15, MDE used one test to measure program effectiveness—the Kindergarten Readiness Assessment. Since then, MDE improved its ability to measure program effectiveness by adding two additional tests—the Brigance III and the Classroom Assessment Scoring System (CLASS). These tests are used nationwide and are better suited for measuring the success of the program.

Inadequate Method for Scoring Sites and Collaboratives

MDE's method for evaluating collaboratives and sites results in a "rate of readiness" score, which can lead to probation and ineligibility for program funds. This "rate of readiness" calculation fails to adequately measure collaborative and site performance. For example, MDE's scoring method for the Kindergarten Readiness Assessment fails to account for student starting scores; thus, a collaborative might do well or poorly based on the natural abilities of the students and not on the contributions of the collaboratives themselves. Also, this scoring method incorporates several arbitrary values.

Use of Curriculum That is Not Evidence-based or Research-based

Beginning in the 2019-20 school year, MDE now requires that all collaboratives use the Opening the World of Learning (OWL) curriculum. However, this curriculum does not qualify as evidenced-based according to MISS. CODE ANN. Section 27-103-159 (1) (a) (1972), and there is not sufficient research to demonstrate its effectiveness in improving student learning.

PEER's Independent Evaluation of Program Effectiveness

Evaluation of Kindergarten Readiness Assessment Scores

The Kindergarten Readiness Assessment (KRA) is a literacy test that classifies children into one of four levels. PEER's evaluation of KRA scores found that six collaborative sites (9%) performed significantly better than the comparison group (i.e., non-collaborative prekindergarten students), and two collaborative sites (3%) performed significantly worse. The remaining 57 sites (88%) did not perform significantly better or worse than the comparison group.

See pages 26-27 for more information on the KRA scores.

Evaluation of Brigance III Scores

The Brigance III is a test given to students to measure their growth in various areas (e.g., motor skills, mathematical skills) over the course of a school year. PEER's evaluation of Brigance III scores found that average scores improved for the majority of sites from the pre-test to the post-test in the 2018-19 school year. Twenty-five sites (39%) improved by a statistically significant amount (i.e., to a degree distinguishable from chance). No site experienced a decrease that was statistically significant. A closer look at those sites with statistically significant improvements could provide valuable insight into what factors might be contributing to these positive results.

See pages 28-30 for more information on the Brigance III scores.

PEER found that certain sites with significant results should be investigated by MDE, as these results are highly unlikely. Conversely, two sites exhibited high, but realistic performance on all assessments (i.e., the KRA, the Brigance III, and the CLASS); these sites could be consulted to determine how they are achieving these positive results.

Evaluation of Classroom Assessment Scoring System (CLASS) Scores

The final test MDE uses in its effectiveness evaluation is the Classroom Assessment Scoring System (CLASS), which is divided into three domains:

- *Emotional Support* - Mississippi's collaborative classrooms were highly successful in the Emotional Support domain of the CLASS, with fifty percent of classrooms scoring above the 90th percentile nationally. The Emotional Support domain measures the social and emotional aspects of prekindergarten classrooms, and some research suggests that the provision of a stable daytime environment (such as classrooms scoring high on this domain) is an important benefit of prekindergarten. Although there was widespread success in this domain, nine classrooms scored 5.5 or below on the 7-point scale. MDE should assess these classrooms to determine the reasons for these low scores.

- *Classroom Organization* - Mississippi's collaborative classrooms were close to the national norms on the Classroom Organization domain of the CLASS, which measures the classroom's management of students. MDE should take a closer look at the twenty classrooms that scored on the bottom end of the distribution to determine why their scores are so low in this domain.

- *Instructional Support* - Mississippi's collaborative classrooms performed worse than the national norms on the Instructional Support domain of the CLASS, which measures the extent to which teachers implement the curriculum to effectively promote cognitive and language development. No classrooms performed in the high scoring range for instructional support, and 25 classrooms (20%) performed below the 10th percentile nationally. MDE should focus on assessing the reasons for low performance on the instructional support domain of the CLASS.

See pages 30-34 for more information on the CLASS scores.

2019 Report on the Early Learning Collaborative Act of 2013: An Evaluation of the Operation and Effectiveness of the Program

Introduction

Authority

MISS. CODE ANN. Section 37-21-51 (1972) directs the Mississippi Department of Education (MDE) to implement a voluntary prekindergarten program in the state. According to Section 37-21-51 (3) (g), MDE "shall make an annual report to the Legislature and the Governor regarding the effectiveness of the program." Furthermore, this CODE Section directs the PEER Committee to:

> *…review those reports and other program data and submit an independent evaluation of program operation and effectiveness to the Legislature and the Governor on or before October 1 of the calendar year before the beginning of the next phased-in period of funding.*

The Committee acted in accordance with MISS. CODE ANN. Section 5-3-51 et seq. (1972).

Purpose and Scope

In its FY 2021 budget request, the Mississippi Department of Education is seeking an additional $3.2 million for the state's prekindergarten program, bringing the total program funding to $9.9 million annually. Thus, the Legislature will need information for the 2020 Regular Session to determine whether to expand the program by appropriating the requested additional funds.

The purpose of this report is to present an independent evaluation of the operation and effectiveness of Mississippi's prekindergarten program established under the Early Learning Collaborative Act that will be useful for decisionmakers when considering the program's future funding.

This is the second evaluation of the prekindergarten conducted by PEER, with the first evaluation conducted in 2015 (*The Early Learning Collaborative Act of 2013: Evaluation of the Operations and Effectiveness of the Program*, PEER Report #600, November 17, 2015). The 2015 report found that there was significant room for improvement in the program's effectiveness based on student test data. For example, PEER determined that the average performance

of students in non-collaborative publicly funded prekindergarten programs was better than the average performance of students in the collaboratives by a statistically significant amount. This report provides an updated evaluation of the program's effectiveness, as PEER analyzed results from multiple tools MDE currently uses to assess the program.

This report will address:

- the creation and purpose of the prekindergarten program;
- the statutory requirements of the prekindergarten program;
- the FY 2017-19 funding appropriated to the prekindergarten program and how the funds have been disbursed;
- MDE's evaluation of collaboratives and sites;
- the status of prekindergarten programs nationally and in Mississippi;
- a summary of the research on the impact of public prekindergarten programs; and,
- PEER's independent evaluation of program effectiveness.

Exhibit 1, page 4, contains the statutory definitions of terms used in this report in reference to the prekindergarten program.

While other public (e.g., Head Start) and private prekindergarten programs exist in Mississippi, this evaluation addresses the effectiveness of only those public prekindergarten programs implemented as part of the Early Learning Collaborative Act.

Method

In conducting this review, PEER:

- reviewed relevant sections of state law;
- interviewed selected staff of MDE;
- reviewed MDE's self-reported information in annual reports[1] from 2016 through 2019, as required by MISS. CODE ANN. Section 37-21-51 (3) (g) (1972), plus additional information requested by PEER;
- requested and analyzed the following test data from MDE:
  - pre-test and post-test scaled scores for each student who took the Kindergarten Readiness Assessment and/or the Brigance III in school year 2018-19 as part of the Early Learning Collaborative Act;
  - pre-test and post-test scaled scores for each student enrolled in a four-year-old prekindergarten program in Mississippi who took the Kindergarten Readiness

[1] PEER did not independently verify all of the information provided in MDE's annual reports.

Assessment but was not a part of the Early Learning Collaborative Act in school year 2018-19;

- pre-test and post-test scaled scores for each student who took the Brigance III in school year 2018-19 as part of the Early Learning Collaborative Act; and,
- scores on the Classroom Assessment Scoring System (CLASS) for each classroom participating in the Early Learning Collaborative Act in school year 2018-19. PEER also reviewed national summary data on the CLASS.

Exhibit 1: Statutory Definitions of Terms Related to the Early Learning Collaborative Act of 2013

Term	Definition
Preschool or prekindergarten children	Children who have not entered kindergarten but will have obtained four years of age on or before September 1 of a school year.
Early Learning Collaborative	A district or countywide council that writes and submits an application to participate in the voluntary prekindergarten program. An early learning collaborative is comprised, at a minimum, of a public school district and/or a local Head Start affiliate if in existence, private or parochial schools, or one or more licensed child care centers. Agencies or other organizations that work with young children and their families may also participate in the collaborative to provide resources and coordination even if those agencies or organizations are not prekindergarten providers.
Prekindergarten Provider	A public, private or parochial school, licensed child care center or Head Start center that serves prekindergarten children and participates in the voluntary prekindergarten program.
Lead Partner	A public school district or other nonprofit entity with the instructional expertise and operational capacity to manage the early learning collaborative's prekindergarten program as described in the collaborative's approved application for funds. The lead partner serves as the fiscal agent for the collaborative and shall disburse awarded funds in accordance with the collaborative's approved application. The lead partner must facilitate a professional learning community for the teachers in the prekindergarten program and lead the collaborative. The lead partner ensures that the collaborative adopts and implements curriculum and assessments that align with the comprehensive early learning standards. The public school district shall be the lead partner if no other qualifying lead partner is selected.
Comprehensive Early Learning Standards	Standards adopted by the State Board of Education that address the highest level of fundamental domains of early learning to include, but not be limited to, physical well-being and motor development, social/emotional development, approaches toward learning, language development and cognition and general knowledge. The comprehensive early learning standards shall also include standards for emergent literacy skills, including oral communication, knowledge of print and letters, phonological and phonemic awareness, and vocabulary and comprehension development.
Research-based Curriculum	A curriculum that has had at least one rigorous controlled evaluation demonstrating effectiveness and does not have an equivalent or more probative body of evaluation demonstrating its ineffectiveness.
Evidence-based Curriculum	A curriculum that has had multiple-site randomized controlled trials across heterogeneous populations demonstrating that it is effective for the population and that does not have an equivalent or more probative body of rigorous evaluation demonstrating its ineffectiveness.

SOURCE: MISS. CODE ANN. Section 37-21-51 (1) (1972) and MISS. CODE ANN. Section 27-103-159 (1) (1972).

Background

This chapter addresses:

- the creation and purpose of the prekindergarten program of the Early Learning Collaborative Act;
- the statutory requirements of the prekindergarten program;
- the funding appropriated to the prekindergarten program and how the funds have been disbursed;
- MDE's evaluation of collaboratives and sites;
- the status of prekindergarten programs nationally and in Mississippi; and,
- a summary of the research on the impact of public prekindergarten programs.

Creation and Purpose of the Prekindergarten Program

In its 2013 Regular Session, the Legislature enacted the "Early Learning Collaborative Act of 2013," which directs the Mississippi Department of Education to implement a voluntary prekindergarten program in the state on a phased-in basis. The prekindergarten program's purpose is to help ensure that all children have access to quality early childhood education and development services.

MISS. CODE ANN. Section 37-21-51 (2) (1972) outlines the following findings of the Legislature with regard to ensuring that all children have access to quality early childhood education and development services:

> *(a) Parents have the primary duty to educate their young preschool children;*
>
> *(b) The State of Mississippi can assist and educate parents in their role as the primary caregivers and educators of young preschool children;*
>
> *(c) There is a need to explore innovative approaches and strategies for aiding parents and families in the education and development of young preschool children; and,*
>
> *(d) There exists a patchwork of prekindergarten entities but no coordination of services and there needs to be a coordination of these services.*

Based upon these findings, the Legislature authorized and directed MDE to implement a prekindergarten program in the state on a phased-in basis. MISS. CODE ANN. Section 37-21-51 (3) (1972), also known as the Early Learning Collaborative Act of 2013, states:

> *Effective with the 2013-2014 school year, the Mississippi State Department of Education shall establish a voluntary prekindergarten program, which shall be a collaboration among the entities*

providing prekindergarten programs including Head Start, licensed child care facilities and licensed public, parochial and private school prekindergarten programs. This program shall be implemented no later than the 2014-2015 school year.

In the 2013-14 school year, eleven early learning collaboratives began implementing a prekindergarten program. Since then, an additional nine collaboratives have implemented a prekindergarten program.

Statutory Requirements of the Prekindergarten Program

MISS. CODE ANN. Section 37-21-51 (1972) outlines MDE's responsibilities in administering the prekindergarten program, eligibility criteria for the awarding of funds to collaboratives, and funding requirements of the program.

MISS. CODE ANN. Section 37-21-51 (3) (b) (1972) states that the prekindergarten program "shall be a collaboration among the entities providing prekindergarten programs including Head Start, licensed child care facilities and licensed public, parochial and private school prekindergarten programs."

MDE's Responsibilities

MISS. CODE ANN. Section 37-21-51 (3) (1972) designates MDE as the entity responsible for administering the implementation, monitoring, and evaluation of the voluntary prekindergarten program, including awards and the application process. Specifically, MDE's statutory responsibilities include:

- establishing a rigorous and transparent application process for the awarding of funds;
- establishing monitoring policies and procedures that, at minimum, include at least one site visit per year;
- providing technical assistance to collaboratives and their providers to improve the quality of prekindergarten programs;
- evaluating the effectiveness of each early childhood collaborative and each prekindergarten provider;
- ensuring that collaboratives provide each parent enrolling a child in the program with a profile of every prekindergarten provider participating in the collaborative's geographic catchment area;
- adopting a minimum rate of readiness that each prekindergarten provider must meet in order to remain eligible for program funds (MDE must recognize each child's unique pattern of development when adopting a rate); and,
- making an annual report to the Legislature and the Governor regarding the effectiveness of the program.

Eligibility Criteria for the Program

Per MISS. CODE ANN. Section 37-21-51 (3) (b) (1972), eligible entities may submit applications for program funds for the following purposes:

- to defray the cost of additional and/or more qualified teaching staff, appropriate educational materials and equipment and to improve the quality of educational experiences offered to four-year-old children in early care and education programs; and/or,
- to extend developmentally appropriate education services at such programs currently serving four-year-old children to include practices of high quality instruction; and,
- to administer, implement, monitor and evaluate the programs; and,
- to defray the cost of professional development and age-appropriate child assessment.

According to MISS. CODE ANN. Section 37-21-51 (3) (d) (1972), prekindergarten program funds must be awarded to early childhood collaboratives whose proposed programs meet certain program criteria (e. g., meet certain teacher-child ratios).

Requirements for Funding

According to MISS. CODE ANN. Section 37-21-51 (3) (h) (iii) (1972), funding must be provided to collaboratives on a basis of $2,150 per student in a full-day program and $1,075 per student in a half-day program. MDE cannot reserve more than 5% of the appropriation for administrative costs. Funds may be carried over to the next year if they are not used.

Collaboratives must match state funds on a 1:1 basis. Local matching funds can include local tax dollars; federal dollars as allowed; parent tuition; philanthropic contributions; or in-kind donations of facilities, equipment, and services required as part of the program (such as food service or health screenings). The Early Learning Collaborative Act of 2013 provided for a state income tax credit for contributions paid to approved providers or collaboratives not to exceed $1 million by any individual, corporation, or other entity.

FY 2017-19 Program Funding

Mississippi's prekindergarten program received legislative appropriations of $14.5 million from FY 2017-19. This funding assisted fourteen early learning collaboratives in implementing prekindergarten programs that served approximately 2,220 students in the 2018-19 school year.

As shown in Exhibit 2, page 8, the Legislature appropriated $4 million in FY 2017, $4 million in FY 2018, and $6,529,634 in FY 2019 ($14.5 million total) for the prekindergarten program. MDE distributed approximately $13.9 million to fourteen early learning

collaboratives and retained $640,992.83 for administrative costs from FY 2017 through FY 2019.

Exhibit 2: Prekindergarten General Fund Appropriation Amounts and Distribution of Funds, FY 2017-19

Year	Legislative Appropriation Amount	Funds Distributed to Collaboratives	MDE's Administrative Costs
FY 2017	$4 million	$3,833,881.19	$166,118.81
FY 2018	$4 million	$3,802,598.49	$197,401.51
FY 2019	$6,529,634*	$6,252,161.49	$277,472.51
TOTAL	**$14,529,634**	**$13,888,641.17**	**$640,992.83**

*In the 2018 Regular Session, the Legislature appropriated a one-time allocation of $2,529,634 to the Early Learning Collaborative Act from funds received by the Office of the Attorney General.

As provided in MISS. CODE ANN. Section 37-21-51 (3) (h) (v) (1972), funds remaining after awards to early learning collaboratives and administrative costs may be carried over in the following year. According to MDE, administrative and carryover funds have been used for two Early Learning Collaborative coordinator positions, a part-time early childhood contractual employee, travel, and commodities for operations and professional development.

SOURCE: MDE.

Selection Process and Collaboratives that Received Funding in FY 2019

In June 2019, the Board of Education approved continued funding for thirteen of the fourteen collaboratives that were already serving students (i.e., Cohort I and Cohort II) in FYs 2017-19. In December 2018, the board approved five new collaboratives (i.e., Cohort III), which began serving students in the fall of 2019.

According to MISS. CODE ANN. Section 37-21-51 (3) (c) (1972), MDE is responsible for administering the implementation, monitoring, and evaluation of the prekindergarten program, including awards and the application process. Exhibit 3, page 9, shows the application and award periods for cohorts and collaboratives.

Exhibit 3: Application and Award Periods for Cohorts and Collaboratives

Cohort[2]	Number of Collaboratives	Year of Initial Application and Award (RFP Cycle)	Year(s) Approved for Continued Funding	Award Period(s)
I	11*	2013-14		January 1, 2014-June 30, 2016
			2016	July 1, 2016-June 30, 2019
			2019	July 1, 2019-June 30, 2024
II	4*	2016-17		January 1, 2017-June 30, 2019
			2019	July 1, 2019-June 30, 2024
III	5	2018-19	N/A	January 1, 2019-June 30, 2022

*Two collaboratives (one from Cohort I and one from Cohort II) did not reapply for the continuation funding grants and are no longer active.

SOURCE: PEER analysis of MDE documentation.

Application Process

MDE has completed three RFP cycles (in 2013-14, 2016-17, and 2018-19) for Early Learning Collaborative awards. MDE's evaluation of collaborative proposals includes the following three stages:

- *Stage 1* of the evaluation involves MDE reviewing the proposals for eligibility. Examples of criteria for eligibility include the designation of a Lead Partner that must serve four-year-old children in a classroom by providing direct instruction, and the Lead Partner must have the instructional expertise and operational capacity to implement the collaborative's proposed plan.
- *Stage 2* of the evaluation involves a team of reviewers who score each eligible proposal based on a set of statutory criteria related to program requirements. For example, the collaborative must meet teacher/child ratios of 1:10 with a maximum of twenty and minimum of five children per classroom. Also, collaboratives must participate in state assessments.
- *Stage 3* of the evaluation involves MDE conducting interviews with the collaboratives that had the highest scores on their proposals. Based on the combined scores from the proposal and interviews from all evaluation team members, MDE makes recommendations to the Board of Education for approval.

[2] A cohort represents a group of early learning collaboratives that applied during the same RFP cycle.

The number of collaboratives that applied for funds by submitting proposals to MDE were:

- 30 collaboratives in the first RFP cycle (2013-14);
- 10 collaboratives in the second RFP cycle (2016-17); and,
- nine in the third application cycle (2018-19).

MDE's three RFP cycles resulted in approved funding for twenty collaboratives within three cohort groups. Two collaboratives chose not to apply for continued funding—the DeSoto County Early Learning Collaborative in 2016, and the Canton Early Learning Collaborative in 2019. Thus, eighteen collaboratives received funding for the 2019-20 school year.

Applications for Continued Funding

In order to receive continued funding, collaboratives must submit a continuation application, which requires collaboratives to provide information related to the scope of work and the budget, and to complete required forms (i.e., the Partner Identification Form, which must be signed by each partner in the collaborative).

According to MDE's continuation application, in order to continue to receive funds for prekindergarten program, programs in Head Start centers, licensed childcare facilities, public, parochial or private schools must form and maintain a stakeholder council called an Early Learning Collaborative involving at least two of the aforementioned entities.

Ten collaboratives from Cohort I applied for and received continued funding for FYs 2017-19, and again for FYs 2020-24. Three collaboratives from Cohort II applied for and received continued funding for FYs 2020-24.

Approved Funding for Cohort III

In the 2018 Regular Session, the Legislature appropriated a one-time $2.5 million General Fund allocation to the Early Learning Collaborative Act. Given this allocation, MDE released an RFP to add new collaboratives. In December 2018, the State Board of Education approved five additional early learning collaboratives, which comprise Cohort III. This cohort began serving students in the 2019-20 school year.

Exhibit 4, page 11, lists the collaboratives operating in FYs 2017-19, along with their associated funding, number of sites, classrooms, and number of students served.

Exhibit 4: FY 2017-19 Funding, Number of Sites, Classrooms and Students Served, by Cohort and Collaborative

Collaborative	FY 2017	FY 2018	FY 2019	Total Funding*	FY 19 Number of Sites (Classrooms)	FY 19 Number of Students
Clarke County Early Learning Partnership	$172,000	$172,000	$172,000	$516,000	2 (5)	97
Coahoma County Pre-K Collaborative Initiative	$327,554	$327,554	$327,554	$982,662	9 (12)	200
Corinth-Alcorn-Prentiss Early Learning Collaborative	$519,225	$519,225	$519,225	$1,557,675	7 (15)	272
Monroe County Early Learning Collaborative**	$427,850	$427,850	$464,400	$1,320,100	8 (12)	206
Lamar County Early Learning Collaborative	$215,000	$215,000	$215,000	$645,000	8 (10)	200
McComb Community Collaborative for Early Learning Success	$462,250	$462,250	$462,250	$1,386,750	5 (12)	218
Petal Early Learning Collaborative	$107,000	$107,000	$107,000	$321,000	2 (4)	79
Picayune Early Learning Collaborative	$43,000	$43,000	$43,000	$129,000	2 (2)	39
Sunflower County Early Learning Collaborative	$316,050	$316,050	$316,050	$948,150	6 (15)	166
Tallahatchie Early Learning Alliance	$344,000	$344,000	$344,000	$1,032,000	4 (11)	192
Cohort I Total	*$2,933,929*	*$2,933,929*	*$2,970,479*	*$8,838,337*	*53 (98)*	*1,669*
Canton Early Learning Collaborative	$247,250	$247,250	$247,250	$741,750	3 (6)	115
Greenwood-Leflore County Early Learning Collaborative	$247,250	$247,250	$247,250	$741,750	2 (10)	176
Grenada Early Learning Collaborative	$247,250	$247,250	$247,250	$741,750	2 (8)	136
Starkville-Oktibbeha Early Learning Collaborative	$247,250	$247,250	$247,250	$741,750	2 (7)	124
Cohort II Total	*$989,000*	*$989,000*	*$989,000*	*$2,967,000*	*9 (31)*	*551*
TOTAL	**$3,922,929**	**$3,922,929**	**$3,959,479**	**$11,805,337**	**62 (129)**	**2,220**

*"Total Funding" represents the not-to-exceed amount for each collaborative.

**Formerly the Gilmore Early Learning Initiative Collaborative.

SOURCE: MDE.

MDE's Evaluation of Collaborative and Sites

MDE evaluates the effectiveness of collaboratives and sites in two ways: by determining their "rate of readiness" as evidenced by scores on three assessment tools, and by monitoring each site to ensure compliance with program and fiscal requirements. Based on rate of readiness scores for the 2018-19 school year, of the 62 sites, 59 were rated successful and three were placed on probation.

According to MISS. CODE ANN. Section 37-21-51 (3) (c) (iv) (1972):

> *The department will evaluate the effectiveness of each early childhood collaborative and each prekindergarten provider. If the State Department of Education adopts a statewide kindergarten screening that assesses the readiness of each student for kindergarten, the State Department of Education shall adopt a minimum rate of readiness that each prekindergarten provider must meet in order to remain eligible for prekindergarten program funds.*

In March 2016, the State Board of Education approved an evaluation to measure the effectiveness of each collaborative and site. (Each site consists of multiple classrooms.) The evaluation consists of two parts—the rate of readiness and monitoring.

Rate of Readiness

MDE measures the rate of readiness for each site and collaborative using three assessment tools: the Kindergarten Readiness Assessment, the Brigance III, and the Classroom Assessment Scoring System (CLASS).

MDE measures the rate of readiness by assigning each site within each collaborative a total score, which includes points from three components:

- students' scores on the Kindergarten Readiness Assessment,[3] a computer-based adaptive assessment that classifies student performance into one of four levels: Early Emergent Reader, Late Emergent Reader, Transitional Reader, and Probable Reader. For the 2018-19 school year, the Kindergarten Readiness Assessment was administered to students in all public prekindergarten programs in the state—collaboratives and non-collaboratives;
- students' scores on the Brigance III[4], an early learning observational screener administered to students by their classroom teachers. For the 2018-19 school year, the Brigance III was only administered to students in collaboratives. In July 2019, MDE issued an RFP to enter into a contract with a company to administer the Brigance III to all public prekindergarten programs in the state—collaboratives and non-collaboratives. MDE anticipates the

[3] The Kindergarten Readiness Assessment was developed by Renaissance Learning to measure literacy, and is administered to prekindergarten students in Mississippi only.

[4] The Brigance III was developed by Curriculum Associates and is a widely used screening tool for students in prekindergarten and compares each child's results with the performance of other test takers.

contract start date to be January 9, 2020, and to continue through June 2024; and

- scores from the Classroom Assessment Scoring System (CLASS)[5], which measures classroom quality by assessing how well the classroom environment and teaching practices support students' learning. CLASS is administered by MDE Certified CLASS Observers, who rate various factors on a 7-point scale, from low to high. CLASS assesses teachers through three domains:
    - *Emotional Support* –includes positive climate, negative climate, teacher sensitivity, and regard for student perspective;
    - *Classroom Organization* – includes behavior management, productivity, and instructional learning formats; and
    - *Instructional Support* – includes concept development, quality of feedback, and language modeling.

  Each classroom receives a four-cycle observation, with each cycle lasting 15-20 minutes.

Rate of Readiness Scoring and Results

MDE placed three sites on probation due to their rate of readiness scores. These sites must achieve a successful rating after one year to receive continued funding.

MDE scores each site on a 100-point scale. The Kindergarten Readiness Assessment scores account for 25 points; the Brigance III scores account for 25 points; and the CLASS ratings account for 50 points.

MDE assigns each site one of the following overall rates of readiness:

- *Successful* – site score of 70+ points;
- *Probation* – site score of 69 points and below. The first year in this category results in a one-year probationary period, in which the site must score 70+ points to receive continued funding;
- *Non-eligible* – site score of 69 and below and has been on probation for one year.

For an overall score breakdown by assessment tool, refer to Exhibit 5, page 21.

Based on rate of readiness scores for the 2018-19 school year, 59 sites were rated successful and three sites were placed on probation and must achieve a successful rating within one year to receive continued funding. Two of the three sites on probation are

[5] The Classroom Assessment Scoring System was developed at the Curry School Center for Advanced Study of Teaching and Learning at the University of Virginia and is widely used to assess classroom quality in prekindergarten-12 classrooms.

in the Canton Early Learning Collaborative, and both low scores resulted from scoring "0" points on the Brigance III. The other site on probation is in the McComb Community Collaborative for Early Learning Success, and its low score resulted from scoring only 15 out of 50 points possible on the CLASS.

No sites were placed on probation due to students' Kindergarten Readiness Assessment scores. Those with the lowest scores on the Kindergarten Readiness Assessment scored high enough on the other two assessment tools (i.e., the Brigance III and the CLASS) to compensate for the lower test scores.

Monitoring

MDE monitors each site's compliance with the Early Learning Collaborative Act and other MDE requirements by reviewing documentation submitted annually by each site and by conducting annual site visits.

For monitoring, MDE uses an Early Learning Collaborative Monitoring Tool to assess compliance with program and fiscal requirements. The monitoring tool primarily consists of administrative items, including teacher credentials, square footage of classrooms, lesson planning procedures, etc. Sites must submit documentation to MDE demonstrating compliance with each program component. Also, MDE representatives visit each site at least once per year between November and May.

MDE completed 14 site visits from April to May 2019, which consisted of desk audits and meetings with staff. These monitoring activities resulted in each collaborative meeting, partially meeting, or not meeting elements of the Early Learning Collaborative Act or other MDE requirements. For any areas in which sites did not achieve compliance, they were required to create a plan with action items and due dates for compliance.

Status of Prekindergarten Programs

According to the National Institute for Early Education Research (NIEER), 1.3 million four-year-olds (33%) were enrolled in public prekindergarten programs in 44 states in 2017-18. Mississippi enrolled five percent of the state's four-year-olds in its early learning collaboratives; however, according to MDE, when combined with school district-funded programs, approximately 18% of the state's four-year-olds were enrolled in public prekindergarten programs in 2017-18. NIEER's 2018 annual report noted that Mississippi met nine of NIEER's ten standards for effective prekindergarten education policies.

The Rutgers Graduate School of Education's National Institute for Early Education Research (NIEER) publishes an annual State of Preschool report, which contains information related to funding, access, and policies of public prekindergarten programs across the nation.

Enrollment

In its most recent annual report for 2018, NIEER found that 1,577,761 students (ages 3-4) were enrolled in 61 public prekindergarten programs in 44 states and the District of Columbia in 2017-18. Nationally,

- 33% of four-year-olds (1.3 million) were enrolled in public prekindergarten;
- 7% of four-year-olds were enrolled in Head Start;
- 3% of four-year-olds were enrolled in special education; and,
- the remaining 57% were not enrolled in public prekindergarten.

The report showed that enrollment in public prekindergarten in 2017-18 varied widely among states. Six states had no prekindergarten program—Idaho, Indiana, New Hampshire, South Dakota, Utah, and Wyoming. The majority of states enrolled between one and 50 percent of four-year olds (e.g., Mississippi, Alabama, Arkansas, Louisiana, and Tennessee). Mississippi enrolled five percent. Eight states and the District of Columbia enrolled between 51 and 85 percent of four-year-olds (e.g., Florida, Georgia, and West Virginia).

According to MDE, when combined with district-funded prekindergarten programs, approximately 18% of Mississippi's four-year-olds attended public prekindergarten programs in 2017-18. Many of the state's children also attend private prekindergarten programs; however, there is no single and complete source from which to obtain data on enrollment in these programs.

Funding

Total state prekindergarten spending reached over $8 billion nationwide. Average state spending per child was $5,175 although this amount varied widely across the states, with a range of $777 to $17,545. Two states and the District of Columbia spent over $10,000 per child, while North Dakota spent the lowest of all states at $777 per child. Eight states spent less than $3,000 in state funds per child, including Mississippi, which spent $2,161 per child.

Quality Standards Benchmarks

NIEER set ten quality standards benchmarks as minimums for effective prekindergarten education and measured state policies against those benchmarks. The benchmarks include:

- Early Learning and Development Standards (ELDS);
- curriculum supports;
- teacher degree;
- teacher specialized training;
- assistant teacher degree;

- staff professional development;
- maximum class size (20);
- staff-child ratio (1:10);
- screenings and referrals; and
- Continuous Quality Improvement System (CQIS).

In its annual report, NIEER determined that three states met all ten standards and six states met nine standards (including Mississippi). Twelve states met less than half of the quality standards benchmarks.

Professional Development Benchmark

Mississippi did not meet the standard related to staff professional development, which requires at least 15 hours of professional development per year for teachers and assistants, individual professional development plans, and coaching or similar classroom-embedded support. Mississippi did not provide professional development through coaching or similar ongoing classroom-embedded support and therefore did not meet the NIEER standard. However, MDE expects to meet all ten standards for the 2019 annual report (to be released in 2020) because it hired a team of early childhood education coaches at the beginning of the 2018-19 school year.

Research on Impact of Prekindergarten Programs

Research evidence suggests that public prekindergarten programs have a positive short-term impact on children's cognitive abilities and their readiness for kindergarten. However, there is not sufficient evidence to conclude that public prekindergarten programs have a positive longer-term impact; two rigorous research studies found that the positive impacts of prekindergarten programs dissipated as children entered elementary school. In light of these studies, it is essential for the state to develop a research strategy to demonstrate prekindergarten program impacts, and it is incumbent upon MDE to identify the best approaches to use in early learning collaborative classrooms that will lead to positive short and long-term impacts.

According to a 2017 report by the Brookings Institution and the Duke Center for Child and Family Policy at Duke University, numerous studies show that children who attend public prekindergarten programs are more prepared to enter kindergarten than children who do not attend such programs. In particular, these children show notable improvements in literacy and numeracy skills. A recent rigorous research study published in February 2019 reiterated these positive short-term benefits for students participating in North Carolina's prekindergarten program.[6] However, there is a lack of solid evidence demonstrating long-term impacts of prekindergarten programs, which is partially due to the

[6] *Effects of the North Carolina Pre-Kindergarten Program, Findings through Pre-K of a Small Scale RCT Study*, Frank Porter Graham Child Development Institute at the University of North Carolina, February 2019.

lack of evidence-based research focusing on longer-term effects. A few studies have found that advantages of attending public prekindergarten extend into the elementary school years, but those studies relied on less than ideal research designs.

Two relatively recent studies used strong research designs (i.e., randomized controlled trials)—one in 2010 based on Head Start[7] and one in 2015 on the Tennessee Voluntary Pre-K Program.[8] Both studies found positive effects on student achievement at the end of the prekindergarten year. However, the Head Start study found that those positive effects for four-year-olds dissipated by the end of kindergarten. The Tennessee study also found that positive cognitive effects dissipated in the years after prekindergarten and further results showed a negative effect by the third grade (i.e., the control group scored significantly higher in math and science achievement than the prekindergarten group). It should be noted that the Head Start study also presented findings related to how prekindergarten affected children's social-emotional and health needs in addition to their cognitive needs. Similar to cognitive effects, the study found that, while prekindergarten had positive social-emotional and health impacts on children (e.g., child received dental care), few impacts remained by the end of the first grade.

In light of these studies, it is incumbent upon the state to conduct the necessary research to demonstrate impacts, and for MDE to identify the best approaches to use in early learning collaborative classrooms that will lead to positive short and long-term impacts. Multiple-site random controlled trials are the most effective way to demonstrate impacts, such as the study conducted in Tennessee.

Researchers from those studies suggest that local sites have wide latitude in the implementation of their programs, which results in diverse approaches; therefore, it is likely that a subset of sites are effective but the impact is offset by ineffective sites. Researchers recommend incorporating an evaluation to identify those approaches that are effective, and require the implementation of those approaches across sites.

The 2017 report by the Brookings Institution and the Duke Center for Child and Family Policy at Duke University also notes that all prekindergarten programs are not all equally effective, as multiple factors may contribute to the success of a program. The researchers identified several "good bets" for supporting strong early learning in prekindergarten including curricula that are known to build foundational skills and knowledge, and professional development and coaching, which enables teachers to create organized and engaging classrooms.

This report addresses these factors in the following ways:

- Pages 23-25 describe the curriculum required in the early learning collaboratives and whether it has been found

[7] *Head Start Impact Study Final Report*, U.S. Department of Health and Human Services Administration for Children and Families Office of Planning, Research and Evaluation, January 2010.

[8] *A Randomized Control Trial of a Statewide Voluntary Prekindergarten Program on Children's Skills and Behaviors through Third Grade*, Peabody Research Institute at Vanderbilt University, September 2015.

through rigorous research to have positive effects on student learning;

- Page 16 addresses the state's policies related to professional development and coaching, and MDE's efforts to increase those functions in the state;
- Pages 30-34 describe the Classroom Assessment Scoring System (CLASS), which seeks to assess classroom quality, including classroom organization, instructional support, and emotional support.

Evaluation of Program Operations and Effectiveness

Since program implementation, MDE has improved its ability to measure program effectiveness by adding two valid tests to its evaluation. However, MDE's method for evaluating effectiveness fails to adequately measure collaborative or site success. Also, MDE requires collaboratives to use a curriculum that does not have research evidence to support its effectiveness in improving student learning. PEER's independent evaluation of scores on each of MDE's tests more accurately measures program and site effectiveness and provides valuable diagnostic information for MDE and the collaboratives. For example, PEER's evaluation of Brigance III scores found that 25 sites (39%) improved from the pre-test to the post-test in the 2018-19 school year by a statistically significant amount (i.e., to a degree distinguishable from chance). Also, PEER's evaluation of Classroom Assessment Scoring System (CLASS) scores for the 2018-19 school year found that collaborative classrooms performed extremely well on the Emotional Support domain when compared nationally, performed close to the national norms on the Classroom Organization domain, and performed worse than the national norms on the Instructional Support domain.

According to MISS. CODE ANN. Section 37-21-51 (3) (g) (1972), MDE "shall make an annual report to the Legislature and the Governor regarding the effectiveness of the [prekindergarten] program." MDE's annual reports, required by MISS. CODE ANN. Section 37-21-51 (3) (g) (1972), show that MDE has made some improvements towards demonstrating program effectiveness. Specifically, MDE has improved its ability to measure program effectiveness by adding two valid tests (i.e., the Brigance III and the Classroom Assessment Scoring System), and those tests have national data by which to compare collaborative performance.

However, MDE still has work to do in order to adequately assess the program and improve its effectiveness. In particular,

- MDE's method for evaluating effectiveness (by calculating a "rate of readiness" for each site) fails to adequately measure collaborative or site success, and
- as of the 2019-20 school year, MDE requires collaboratives to use the Opening the World of Learning (OWL) curriculum; however, OWL is not evidence-based or research-based. Thus, there is no assurance that OWL is effective in improving student learning.

Also, as required by MISS. CODE ANN. Section 37-21-51 (3) (g) (1972), PEER conducted an independent evaluation of the program's effectiveness. Pages 25 through 34 contain the results of this evaluation.

Improved Ability to Measure Program Success

In school year 2014-15, MDE used one test to measure program effectiveness—the Kindergarten Readiness Assessment. Since then, MDE improved its ability to measure program effectiveness by adding two additional tests—the Brigance III and the Classroom Assessment Scoring System (CLASS). These tests are used nationwide and are better suited for measuring the success of the program.

In March 2014, MDE adopted a statewide test to assess the readiness of each prekindergarten student for kindergarten—the

Kindergarten Readiness Assessment (KRA). When PEER conducted its first review of the prekindergarten program in 2015, MDE was using the KRA alone to measure program effectiveness. Because this test only measures students' literacy and does not measure other important content areas (e.g., social and emotional development), it was inadequate to assess total program performance.

Since that time, MDE improved its ability to measure program effectiveness by adding two additional tests—the Brigance III and the Classroom Assessment Scoring System (CLASS). In FY 2016, MDE began using the KRA and the CLASS to measure program effectiveness, and in FY 2018, MDE added the Brigance III.[9]

PEER reviewed the technical documentation of the Brigance III and the CLASS and found they are better documented than the KRA, including better evidence of their reliability and validity. Both tests are used nationwide and are more suitable for measuring success of the program. The Brigance III adds value to the assessment of program effectiveness because it measures students' development and growth over time related to social and emotional development, language development, cognition and general knowledge, physical well-being and motor development, and approaches to learning. The CLASS adds value because it measures classroom quality by assessing dimensions of teaching that are linked to student achievement and development.

Inadequate Method for Scoring Sites and Collaboratives

MDE's method for evaluating collaboratives and sites results in a "rate of readiness" score, which can lead to probation and ineligibility for program funds. This "rate of readiness" calculation fails to adequately measure collaborative and site performance.

As described on pages 13-14, MDE measures the rate of readiness for each site and collaborative using the results from three assessment tools: the Kindergarten Readiness Assessment (KRA), the Brigance III Screens, and the Classroom Assessment Scoring System (CLASS). If a site receives a low score, it may be placed on probation for a year, and then deemed ineligible for program funding if the site score has not improved after one year.

MDE's "rate of readiness" calculation aggregates KRA, Brigance, and CLASS scores. The site score breakdown by assessment tool is described in Exhibit 5, page 21.

[9] The Brigance III replaced the Learning Accomplishment Profile, 3rd edition (LAP-3), which was used in FY 2017 only.

Exhibit 5: Site Score Breakdown by Assessment Tool for 2018-19

Assessment	Criteria for Awarding of Points	Points
Kindergarten Readiness Assessment	Percent of children that meet 498 benchmark by the end of the year	0-29% = 0 points 30-49% = 8 points 50-65% = 15 points 66-100% = 25 points
	OR	OR
	Percent of children that demonstrate an average point gain of 98 per site	0-24% = 0 points 25-39% = 8 points 40-49% = 15 points 50-100% = 25 points
Brigance III	Positive student growth from October to May	25 points
CLASS Ranges • Low = 1-2 • Mid = 3-5 • High = 6-7	Average across domains plus performance on Instructional Support (IS) domain if site average is 5.00 or higher	1.00-2.99 = 0 points 3.00-3.99 = 15 points 4.00-4.99 = 30 points 5.00-7.00 & <2.8 IS = 30 points 5.00-7.00 & ≥ 2.8 IS = 50 points

SOURCE: MDE.

The "rate of readiness" calculation has several flaws and fails to adequately measure collaborative performance, as discussed in the following sections.

Issues with MDE's Scoring of the Kindergarten Readiness Assessment

The Kindergarten Readiness Assessment (KRA) measures students' literacy. MDE's scoring method for the KRA fails to account for student starting scores; thus, a collaborative might do well or poorly based on the natural abilities of the students and not on the contributions of the collaboratives themselves. Also, this scoring method incorporates several arbitrary values. The effect of these manipulations is that it introduces bias into the original measurement of student performance.

The Kindergarten Readiness Assessment is a measurement of students' literacy. Students who achieve a score of 498 at the end of the prekindergarten year on the Kindergarten Readiness Assessment are reportedly on track to achieve proficiency on Mississippi's third-grade reading assessment. However, MDE's scoring method regarding the Kindergarten Readiness Assessment has issues. As shown in Exhibit 5, above, MDE assigns points to sites based on the percentage of each site's students achieving a score of 498 or the percentage of students achieving a score gain of 98.

The scoring method for the KRA fails to account for student starting scores; a collaborative might do well or poorly based on

the natural abilities of its students and not on the contributions of the collaboratives themselves. While student assessment is an important part of education, in the current context, the effectiveness of collaboratives, and not students, is at issue.

The 498 score benchmark is evaluatively sound, as it predicts subsequent proficiency on the third-grade reading test; however, MDE added a way for sites to receive points if a certain percentage of children achieved a 98-point gain from fall to spring. The 98-point gain benchmark is not evaluatively sound, as it is simply derived from the average gain from the first year of test administration. Expected gain should either be explicitly related to the specific year's average, or, preferably, related to a predicted rate of gain by individual student age, based on the representative group used to construct the test.

Also, the KRA scoring method incorporates several arbitrary values (i.e., values that are neither derived from the test itself or the distribution of scores of those taking it, nor related to any valuable outcomes external to the test). The percentage ranges and point values used in this calculation are therefore arbitrary.

The net effect of these arbitrary manipulations is that it introduces bias into the original measurement. First, the calculation takes an original test score or change in test score and applies a benchmark, to transform the test score into a measure of percent-compliant. This first transformation eliminates some information about the distribution of original scores. For example, a group of 100 students, 50 with KRA scores of 300 and 50 with KRA scores of 498, is assigned the same number of points as a group of 100 students, 50 with KRA scores of 497 and 50 with KRA scores of 900.

The original test score has now been transformed into a percent-passing score. However, the calculation now divides collaboratives into four arbitrary categories, reducing the informational content of the transformed score by 25 times (from 100 possible states to four). This categorized score now has an arbitrary point value applied to it, which introduces bias into the original measurement because it adds information to the measurement that is not created by the test but created by MDE.

Issues with MDE's Scoring of the Brigance III and the Classroom Assessment Scoring System (CLASS)

The Brigance III is a valid assessment that can be used to measure students' growth[10] over the course of the school year. However, MDE's scoring of the Brigance III does not reflect an actual measure of students' growth. Also, although the Classroom Assessment Scoring System (CLASS) is a valid measurement of classroom quality, MDE's scoring of the CLASS by assigning categories and point values is arbitrary.

[10] See Technical Appendix, page 35, for a description of how Brigance III scores can be used to measure students' growth.

As mentioned previously, the Brigance III and the CLASS are better suited to measure the effectiveness of the prekindergarten program. While the MDE does not manipulate the scores from the Brigance III and the CLASS as much as it does for the KRA scores, PEER found the following issues that affect the MDE's rate of readiness calculation:

- Although MDE's criteria for awarding of points indicates that sites receive points for positive growth from October to May, MDE's definition of student growth on the Brigance III is a percent-above-benchmark measure, and not a measure of growth. Percent-above-proficient benchmarks do not adequately capture change in performance over time; they ignore any individual who does not cross a benchmark, and are capable of ignoring a trend affecting 100% of the population, or reversing a trend affecting all but a single member of the population. Consider the following hypothetical examples:
    - A test has an arbitrary benchmark of 51, and a student can score between one and 100. A population of 1,000 students takes the pre-test and achieves a score of one. On the post-test, all 1,000 students achieve a score of 50, which represents a 49-point growth for each student. That growth is ignored by a percent-above-benchmark measure, which records only 0% success on the pre-test and 0% success on the post-test.
    - On the same test, a different population of 1,000 students score 51 on the pre-test. On the post-test, 999 students score 100 and one scores fifty; 99.9% of the population achieved growth. However, a percent-proficient measure counts this population as having regressed rather than grown.
- The CLASS scores are subject to less manipulation, and are appropriately assigned a higher proportion of the available points in the rate of readiness calculation; however, its categories and point values are still arbitrary and therefore represent impositions by the observer rather than actual measurements.

Use of Curriculum That is Not Evidence-based or Research-based

Beginning in the 2019-20 school year, MDE now requires that all collaboratives use the Opening the World of Learning (OWL) curriculum. However, this curriculum does not qualify as evidenced-based according to MISS. CODE ANN. Section 27-103-159 (1) (a) (1972), and there is not sufficient research to demonstrate its effectiveness in improving student learning.

MDE revised its requirements for collaboratives related to required curricula from the 2013-14 RFP cycle to the 2016-17 RFP cycle. While the 2013-14 RFP required that participating programs use a *research-based curriculum* designed to prepare students for kindergarten, the 2016-17 RFP and the 2018-19 RFP required that

participating programs use an *evidence-based curriculum* that places an emphasis on early literacy, is aligned with MDE's Early Learning Standards for Classrooms Serving Four-Year-Old Children, and contains thematic units of activities and ideas designed so that children can master the standards.

MDE's 2018-19 RFP for new collaboratives (and 2019 continuation application for already existing collaboratives) further states that MDE identified one curriculum that meets the evidence-based requirement—Opening the World of Learning (OWL). Consequently, MDE only approved funding in 2019 for collaboratives using the OWL curriculum beginning in the 2019-20 school year.

Evidence-based Standard Not Met

OWL does not meet the standard for an evidence-based curriculum. Results from the only randomized controlled trial PEER found involving OWL showed that it performed significantly worse than comparison curricula.

MISS. CODE ANN. Section 27-103-159 (1) (a) (1972) requires that an evidence-based program have multiple randomized controlled trials demonstrating effectiveness, as it states:

> *'Evidence-based program' means an intervention program that has had multiple site randomized controlled trials across heterogeneous populations demonstrating that the program is effective for the population and that does not have an equivalent or more probative body of rigorous evaluation demonstrating its ineffectiveness.*

Evidence-based research is the gold standard for demonstrating causal inference regarding program effectiveness.[11]

OWL does not meet the standard for an evidence-based curriculum according to MISS. CODE ANN. Section 27-103-159 (1) (1972) because it is not based on the findings of multiple-site random controlled trials across heterogeneous populations. To the best of PEER's knowledge, the only randomized controlled trial involving OWL resulted in OWL performing significantly worse than its comparison curricula.[12] Of particular note, one of the comparisons was with Creative Curriculum, which has previously been shown to be ineffective.[13] PEER's 2015 report[14] showed that the MDE awarded funding to four collaboratives using the Creative Curriculum in school year 2014-15. According to MDE, seven Head Start centers in five collaboratives used Creative Curriculum in the 2018-19 school year (prior to the OWL requirement).

---

[11] Coalition for Evidence-based Policy. (2003). *Identifying and Implementing Educational Practices Supported by Rigorous Evidence: A User-friendly Guide.* US Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

[12] Kaiser, A., Dickinson, D., Roberts, M., Darrow, C., Freiberg, J., & Hofer, K. (2011). *The Effects of Two Language-Focused Preschool Curricula on Children's Achievement through First Grade.* Society for Research on Educational Effectiveness.

[13] What Works Clearinghouse, https://ies.ed.gov/ncee/wwc/Docs/InterventionReports/wwc_creativecurriculum_030513.pdf.

[14] https://www.peer.ms.gov/Reports/reports/rpt600.pdf

Research-based Standard Not Met

OWL does not meet the standard for a research-based curriculum (a lower standard than evidence-based) because there is no high-quality research to show that OWL is effective. Furthermore, the one high-quality study PEER found involving OWL showed that it was ineffective in improving student learning.

A lower, research-based standard for demonstrating effectiveness is also available in the law. MISS. CODE ANN. Section 27-103-159 (1) (c) (1972) states the following:

> *'Research-based program' means an intervention program that has had at least one (1) rigorous controlled evaluation demonstrating effectiveness and does not have an equivalent or more probative body of evaluations demonstrating its ineffectiveness.*

The research-based standard requires, at minimum, the use of a control group. PEER is aware of one study testing OWL in isolation using a control group, and this study showed no significant effects of OWL as compared to the control group.[15]

There are a number of studies testing OWL without a control group; however, these studies are too low-quality to support causal inferences about program effectiveness. Based on these findings, PEER concludes that there is no high-quality evidence of OWL's effectiveness, and at least some evidence of its ineffectiveness in improving student learning.

If MDE continues to use OWL, it should work towards incorporating a randomized control trial in the state to contribute to the research on OWL.

PEER's Independent Evaluation of Program Effectiveness

As noted previously, MISS. CODE ANN. Section 37-21-51 (3) (g) (1972) directs the PEER Committee to conduct "an independent evaluation of program operation and effectiveness" of the prekindergarten programs funded through the Early Learning Collaborative Act of 2013. PEER notes that the effectiveness evaluation of this report is limited to the performance of the collaboratives that received funding through the Early Learning Collaborative Act and does not address whether prekindergarten is the best investment of Mississippi tax dollars in comparison to other statewide priorities. Decisions regarding future investment of public funds in prekindergarten programs in Mississippi should be based on statistical analysis of the effectiveness of the programs established by the Early Learning Collaborative Act and on other research targeted to our state's specific needs.

PEER conducted the following analyses on data provided by the MDE:

[15] Abdullah-Welsch, Schmidt, Hahn, Tafoya, & Sifuentes. (2009). *Evaluation of the Opening the World of Learning (OWL) Early Literacy Program: Final Report*. WestEd. December 2009.

- collaborative students' scores on the Kindergarten Readiness Assessment compared to non-collaborative students' scores;
- collaborative students' normalized scores on the Brigance III post-test compared to their scores on the pre-test; and,
- collaborative classrooms' scores on the Classroom Assessment Scoring System (CLASS) compared to national scores of Head Start classrooms.

Evaluation of Kindergarten Readiness Assessment Scores

PEER's evaluation of Kindergarten Readiness Assessment (KRA) scores found that six collaborative sites (9%) performed significantly better than the comparison group (i.e., non-collaborative prekindergarten students), and two collaborative sites (3%) performed significantly worse. The remaining 57 sites (88%) did not perform significantly better or worse than the comparison group.[16]

The Kindergarten Readiness Assessment (KRA) is a literacy test that ranks children on a scale of ability and classifies their performance into one of four levels: Early Emergent Reader, Late Emergent Reader, Transitional Reader, and Probable Reader.

Because the focus of this assessment is on the individual child (not the site or classroom), some amount of inference is required to use the KRA to evaluate a collaborative site or classroom. For example, an evaluation should not simply compare pre-test and post-test scores on the KRA and consider this as a measure of the program's effectiveness. Because ability on the test is expected to increase with age, it is expected that the KRA score will increase also, whether or not prekindergarten is making any contribution to that increase.

Exhibit 6, page 27, demonstrates one inference from KRA scores to collaborative prekindergarten classroom performance: median post-test scores, controlling for age and pre-test scores, as compared to the comparison or control group (i.e., non-collaborative prekindergarten students).

Each horizontal bar on the graph's y-axis represents students' scores in a single collaborative site, and the sites are arranged in order from the greatest (at the top of the exhibit) to the least (at the bottom of the exhibit) value. These sites are anonymized in order to avoid any potential Family Educational Rights and Privacy Act (FERPA) violation involving personally identifiable information.

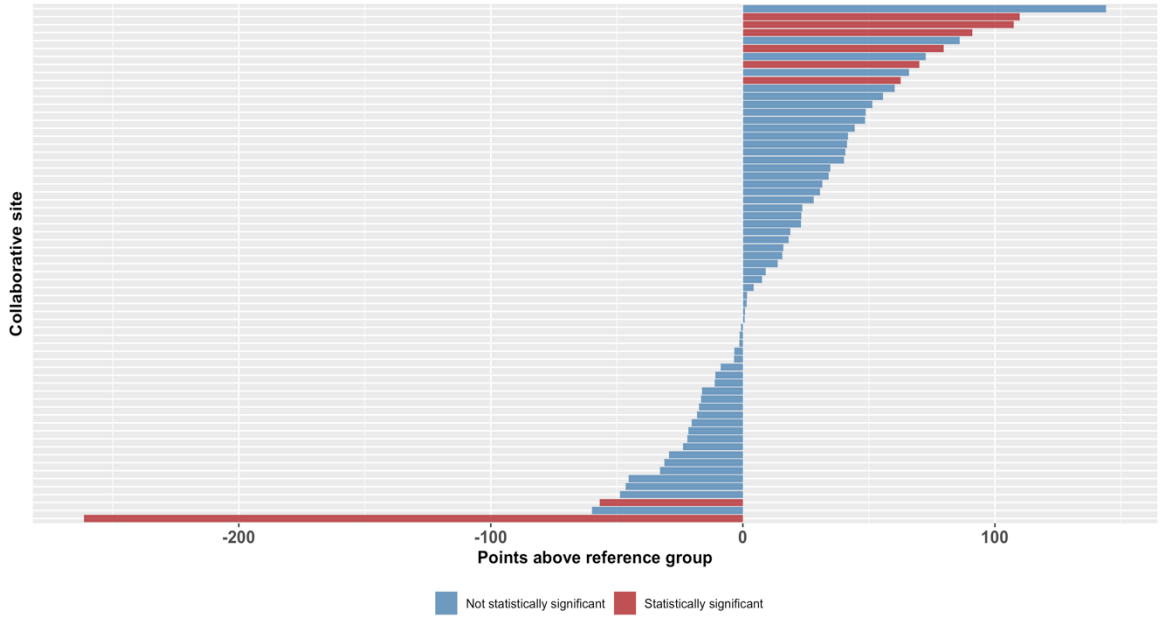
The x-axis on Exhibit 6 represents points on the KRA test. For instance, the median of the lowest performing site was over 261 points lower than the median of the non-collaborative students when controlling for other factors, while the median of the highest performing site was over 144 points higher.

[16] While MDE's annual report indicates a total of 62 sites, the KRA dataset provided by MDE to PEER had 65 different site codes. In particular, the Sunflower County Early Learning Collaborative included more site codes in the dataset provided to PEER than is indicated in MDE's annual report.

Collaborative sites with blue bars on Exhibit 6 were different from the non-collaborative students but not to a statistically significant degree. Collaborative sites with red bars exhibited a statistically significant difference (i.e., to a degree distinguishable from chance) from the non-collaborative students.

As shown in Exhibit 6, below, students in the majority of sites performed better than non-collaborative students. Specifically, 40 sites performed better and 25 performed worse. Most differences were not statistically significant, however. Six sites (9%) achieved a positive statistically significant result, and two sites (3%) had a negative statistically significant result. The remaining 57 sites (88%) did not perform significantly better or worse than the comparison group. The small class sizes of some collaborative sites decrease the power to find significant differences (which explains why some blue bars in Exhibit 6 are close in proximity to red bars but are not statistically significant).

Exhibit 6: Difference in Kindergarten Readiness Assessment Post-test Performance of Students Attending Collaboratives Versus Non-collaboratives by Site



SOURCE: PEER analysis of Kindergarten Readiness Assessment scores.

Evaluation of Kindergarten Readiness Assessment Scores in Conjunction with CLASS and Brigance III Scores

PEER found that certain sites with significant results should be investigated, as these results are highly unlikely. Conversely, two sites exhibited high, but realistic performance on all assessments (i.e., the KRA, the Brigance III, and the CLASS); these sites could be consulted to determine how they are achieving these positive results.

PEER took a closer look at these sites in conjunction with other measures of performance (i.e., the Brigance III and the CLASS) and found three issues that warrant further investigation:

- The bottom of Exhibit 6 shows one strong outlier. PEER found that this site's pre-test scores on the KRA were anomalously high, while its post-test scores dropped to a normal range. These results suggest that further investigation is needed to determine why pre-test results were unusually high.
- Two sites exhibited extremely unlikely progress on the KRA and the Brigance. Students' scores increased by a statistically significant amount from pre-test to post-test. Also, these sites' CLASS scores (which measure classroom quality) were extremely low.

In contrast to these results, two sites had high CLASS scores, meaningful progress on the Brigance III, and positive statistically significant results on the KRA. These sites could possibly be consulted to determine how they are achieving these results.

Evaluation of Brigance III Scores

PEER's evaluation of Brigance III scores found that average scores improved for the majority of sites from the pre-test to the post-test in the 2018-19 school year. Twenty-five sites (39%) improved by a statistically significant amount (i.e., to a degree distinguishable from chance). No site experienced a decrease that was statistically significant. A closer look at those sites with statistically significant improvements could provide valuable insight into what factors might be contributing to these positive results.

The second assessment MDE uses in its effectiveness evaluation is the Brigance III, which is a test given to students up to the first grade to measure their growth over the course of a school year. The Brigance III is a more useful measure to demonstrate program effectiveness than the Kindergarten Readiness Assessment because the Brigance III data allow for meaningful comparisons to national data and across children's' ages. Thus, it is better suited for the pre-post testing MDE uses in its evaluation. Also, the Brigance III covers a wider variety of skills than the KRA, including motor and mathematical skills.

Exhibit 7, page 30, shows the average Brigance scores on the pre-test given before prekindergarten (represented on the x-axis) and the post-test given after prekindergarten (represented on the y-axis) for all collaborative sites. Each dot represents a single site comprised of multiple classrooms. The national average is 100 on both the pre- and post-test.

The horizontal and vertical grey lines at 100 on each axis divide Exhibit 7 into four quadrants. The top left quadrant represents sites worse than average on the pre-test, but better than average on the post-test, and the top right contains sites better than average on both tests. The bottom right contains sites better than average on the pre-test, but worse on the post-test, and the bottom left contains sites worse than average on both tests.

The diagonal grey line is the line of zero change; a site exactly on this line performed no differently, relative to the average, on the pre-test than it did on the post-test. Thus, any site above the diagonal line contained students who, on average, improved after prekindergarten; any site below the diagonal line contained students who performed worse after prekindergarten.

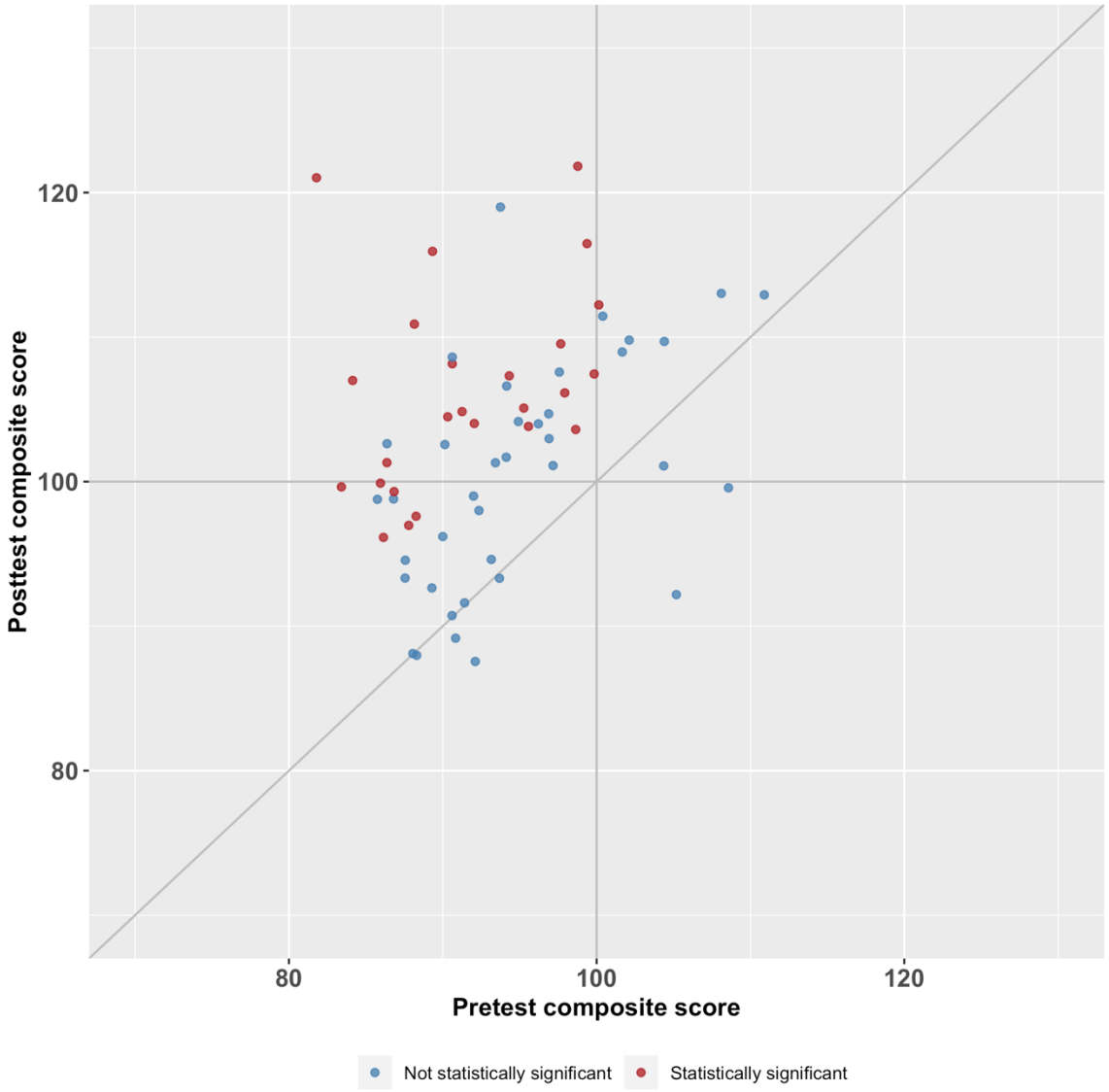
Because it is expected that there is some difference between what is measured and what is actually occurring, it would be unusual if sites lined up exactly on the diagonal, even if no underlying change either way happened during prekindergarten. As such, Exhibit 7 also presents the statistical significance of the difference between each site's pre-test and post-test score distributions. Statistically significant differences are represented in red; differences without statistical significance are represented in blue.

As Exhibit 7 shows, more sites experienced improvement than decline in average scores from the pre-test to the post-test. Twenty-five of the total 64 sites[17] (39%) improved by a statistically significant amount (i.e., to a degree distinguishable from chance). No site experienced a decrease that was statistically significant.

A closer look at those sites with statistically significant improvements could provide valuable insight into what factors might be contributing to these positive results. Those sites with dots farther from the diagonal line in Exhibit 7 are of particular interest, but it may be worth further investigating any outlier school because the small class sizes of some collaborative sites decrease the power to find significant differences (which explains why some blue dots in Exhibit 7 are close in proximity to red dots but are not statistically significant).

[17] While MDE's annual report indicates a total of 62 sites, the Brigance III dataset provided by MDE to PEER had 64 different site codes. In particular, the Sunflower County Early Learning Collaborative included more site codes in the dataset provided to PEER than is indicated in MDE's annual report. Also, two sites did not administer the Brigance III pre-test and post-test, and are therefore excluded from the data.

Exhibit 7: Differences in Average Pre-test and Post-test Scores on the Brigance III by Site for School Year 2018-19



SOURCE: PEER analysis of Brigance III scores.

Evaluation of Classroom Assessment Scoring System (CLASS) Scores

PEER's evaluation of Classroom Assessment Scoring System (CLASS) scores for the 2018-19 school year found that collaborative classrooms performed extremely well on the Emotional Support domain when compared nationally, performed close to the national norms on the Classroom Organization domain, and performed worse than the national norms on the Instructional Support domain.

The final test MDE uses in its effectiveness evaluation is the Classroom Assessment Scoring System (CLASS). This test is different from the Kindergarten Readiness Assessment and the Brigance III because it is directly applicable to prekindergarten classrooms. Furthermore, the U.S. Department of Health and Human Services Office of Head Start collects CLASS data on Head

Start grantees, which provides for valuable comparisons with Mississippi's prekindergarten program. In 2018, the Head Start's Early Learning and Knowledge Center collected and published summary data from 462 Head Start grantees that received CLASS reviews. PEER used this national data in its analysis of Mississippi's CLASS scores.

CLASS scores are divided into three domains:

- Emotional Support, which rates a classroom's social and emotional aspects;
- Classroom Organization, which rates a classroom's management of students; and,
- Instructional Support, which rates a classroom's curriculum implementation and associated support of student cognitive and language development.

Each domain is scored on 7-point scale, with 1-2 being a low score, 3-5 being a middle score, and 6-7 being a high score. Exhibit 8 plots the scores on each of these three domains against approximated national percentiles. (See Technical Appendix for how these were approximated.)

Each of the three graphs in Exhibit 8 plots Mississippi's collaborative prekindergarten classrooms against national percentiles (the y-axis) and the 1-7 CLASS scores (the x-axis) for each domain. The horizontal grey line on each graph represents the national median (i.e., the 50th percentile), while the vertical grey lines represents the low (1-2), middle (3-5), and high (6-7) score areas.

Emotional Support Domain

Mississippi's collaborative classrooms were highly successful in the Emotional Support domain of the CLASS, with fifty percent of classrooms scoring above the 90th percentile nationally. The Emotional Support domain measures the social and emotional aspects of prekindergarten classrooms, and some research suggests that the provision of a stable daytime environment (such as classrooms scoring high on this domain) is an important benefit of prekindergarten.

The top third of Exhibit 8, page 34, shows that Mississippi's collaborative prekindergarten classrooms were extremely successful in the Emotional Support domain of the CLASS in the 2018-19 school year, compared to national data. The Emotional Support domain measures the degree to which teachers establish and promote a positive climate in their classroom through their everyday interactions. Fifty percent of the classrooms (64 out of 128) were in the top half of the high score range (6.5 and above) which is above the 90th percentile nationally. This level of success is highly commendable as it widespread throughout collaborative classrooms.

It is particularly worth noting that, while some rigorous studies suggest that the academic benefits of pre-k are negligible or temporary (see pages 16-18), there is also research suggesting that

the provision of a stable daytime environment for its participants is itself an important benefit of prekindergarten. In other words, some research suggests that the non-academic benefits of prekindergarten are more important than its academic benefits, if any. An important part of providing this stable daytime environment is reflected in the CLASS's Emotional Support domain.

Widespread success is not universal success, of course, and MDE should still pay attention to classrooms that scored well below the mean. Nine classrooms scored 5.5 or below on the 7-point scale and ranged from 1.87 to 5.9 standard deviations below the mean, which is a cause for concern.

Classroom Organization Domain

Mississippi's collaborative classrooms were close to the national norms on the Classroom Organization domain of the CLASS, which measures the classroom's management of students.

The middle graph of Exhibit 8 shows the distribution of prekindergarten collaboratives with regards to the Classroom Organization domain, compared to national data. The classroom organization domain assesses classroom routines and procedures related to the organization and management of children's behavior, time, and attention in the classroom.

On this domain, the collaborative classrooms performed closer to the national average; mean domain scores were below the national average, but median scores were above the national average. This is because the distribution of scores is skewed; there are 20 classrooms below the first percentile on this domain, with the most extreme outlier more than eight standard deviations below the national norm.

With regards to classroom organization, the distribution of scores in Mississippi for the 2108-19 school year is approximately the same as the national distribution, except for the bottom end. The classrooms in the bottom of this distribution are worth a closer look to determine why their scores are so low in this domain.

Instructional Support Domain

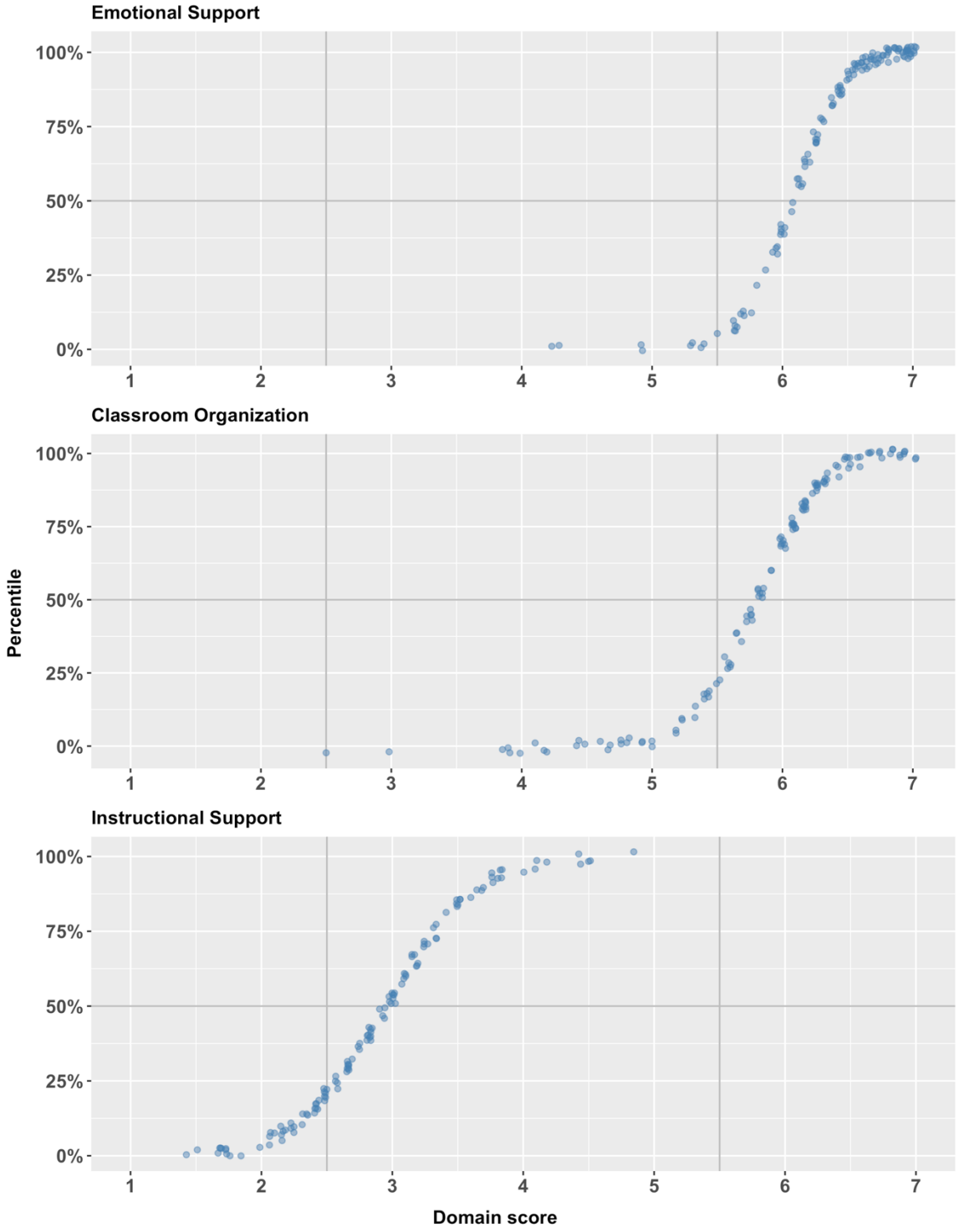
Mississippi's collaborative classrooms performed worse than the national norms on the Instructional Support domain of the CLASS, which measures the extent to which teachers implement the curriculum to effectively promote cognitive and language development. No classrooms performed in the high scoring range for instructional support, and 20% performed below the 10th percentile nationally.

The third graph of Exhibit 8 shows the distribution of collaborative classroom scores with regards to the Instructional Support domain, which assesses the ways in which teachers implement the curriculum to effectively promote cognitive and language development. The fact that the distribution is shifted to the left relative to the other two domains indicates that the collaboratives classrooms' performance in this domain was poorer, in terms of

raw scores, than on the other two domains; however, the national average is also lower on this domain than on the other two. Nonetheless, Mississippi's collaboratives performed worse than the national average in this domain.

Of the 128 classrooms, none performed in the high scoring range, while 43 (34%) were in the low scoring range. Additionally, 73 (57%) were below the national median. Of particular concern, 25 (20%) were below the 10th percentile nationally.

Exhibit 8: Mississippi Collaborative Classroom Scores on the Classroom Assessment Scoring System (CLASS) Compared to National Percentiles by Domain



SOURCE: PEER analysis of CLASS scores.

Technical Appendix

Kindergarten Readiness Assessment

In using the Kindergarten Readiness Assessment (KRA) as part of an evaluation of pre-k, it is important to bear in mind the limited documentation of its psychometric properties. It is an empirical question (though one weighted with a great deal of nonempirical measurement theory) whether any given psychometric test results in interval-scale measurement; even a test constructed according to the Rasch model produces interval measurements only if the underlying data themselves obey the axioms of additive conjoint measurement,[18] which mere model fit does not guarantee. Theoretical mechanisms for testing consistency with these axioms exist, but are rarely applied in the psychometric literature;[19] PEER is not aware of any such tests of the KRA. As such, prudence dictates treating the KRA as an ordinal measurement.

Additionally, the KRA is intended as a measurement of an underlying ability that varies with age; older children are expected to have more developed linguistic skills than younger children, all else being equal, simply by virtue of being older. Ideally, the KRA would include some form of age normalization, in order to ease measurements of growth and comparisons among groups of different ages. The fact that a child's KRA score improves over the course of pre-k is unremarkable given that the child also ages during that time; the relevant question is whether it improves more than we would expect given the aging process. However, no such normalization has been made available to PEER.

PEER is aware of documentation of one function of the KRA: Predicting success on a subsequent standardized test. PEER's 2015 report on Mississippi's pre-k collaboratives employed the KRA benchmark for likely success on the later test to treat the KRA results as binary, such that the only important question was whether a student was over or under that benchmark. While this was a psychometrically conservative choice, it obscures information; for this year's evaluation, a more informative method that still respects the limited documentation available on the KRA was chosen.

Because this analysis was designed for the assessment of a population rather than causal generalization from a sample to a population, individuals missing either a pre-test or a post-test were dropped from the analysis; the final analysis was thus based on 8171 individuals, divided into 65 collaborative schools and a reference distribution consisting of all students outside the collaboratives.

Statistical analysis was conducted using a quantile regression model with bootstrapped standard errors;[20] KRA post-test score was the dependent variable and age at post-test (expressed in days), school, and an interaction between age at pre-test and pre-test score were the independent variables. It is important to note that this model is not intended for causal inference, and it is not intended to generalize beyond the specific population to which it was applied in this report; it is, essentially, simply a multidimensional measure of median school performance relative to the median of a reference distribution. P-values of coefficients were Holm-corrected[21] to control for the testing of multiple hypotheses.

---

[18] Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of mathematical psychology*, *1*(1), 1-27.

[19] Heene, M. (2013). Additive conjoint measurement and the resistance toward falsifiability in psychology. *Frontiers in psychology*, *4*, 246.

[20] Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33-50.

[21] Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70. This method is universally preferable to Bonferroni's method, as established by Aickin

The data plotted in Exhibit 6, page 27, are simply the coefficients of each school in the regression model; they represent the median performance of that school relative to the median performance of non-collaboratives. The data underlying the table, anonymized and presented in descending order of coefficient value, are presented in Table 1 below.

Table 1:

School	Coefficient	Standard error	Holm p-value
1	144.1499751	89.46780678	1
2	109.7909818	14.48933084	2.54E-12
3	107.4951823	27.01682198	0.004331472
4	91.09603757	18.72834468	7.38E-05
5	86.07398667	26.18536849	0.057944074
6	79.67631743	20.78625522	0.007649157
7	72.56156935	46.80083547	1
8	70.07762331	19.17413478	0.015281906
9	65.95823978	20.42184451	0.068407601
10	62.67569161	17.18776432	0.015513699
11	60.2983839	26.33239894	0.992392352
12	55.64312042	18.99218272	0.173471159
13	51.3449032	24.82521067	1
14	48.7631159	21.42085345	1
15	48.4404448	17.00673095	0.220315423
16	44.33054478	14.94592955	0.159725552
17	41.67633437	18.53059433	1
18	41.3583793	22.35088344	1
19	40.62886143	13.69248504	0.159725552
20	40.09912613	21.13085656	1
21	34.69924588	14.46165099	0.756451533
22	34.01647915	24.26801378	1
23	31.53643869	10.43935206	0.136510488
24	30.65400149	14.54938857	1
25	28.13771405	10.81626518	0.446430696
26	23.56762267	8.30115473	0.222249775
27	23.20979121	13.98157236	1

and Gensler (Aickin, M., & Gensler, H. (1996). Adjusting for multiple testing when reporting research results: the Bonferroni vs Holm methods. *American journal of public health*, *86*(5), 726-728.); it is also more conservative than methods of multiple hypothesis correction based on the false discovery rate.

28	23.03718153	16.95847689	1
29	18.90536292	29.14135051	1
30	18.19989094	31.82555026	1
31	16.04905315	11.4429563	1
32	15.63825988	21.82442585	1
33	13.87092807	26.80654637	1
34	8.990511581	9.484873369	1
35	7.648950707	19.4386034	1
36	4.24578948	11.04990667	1
37	1.679415721	17.1061508	1
38	1.452159657	18.09922015	1
39	0.817138691	43.29138077	1
40	0.690767229	12.4172758	1
41	-0.788507402	34.7771757	1
42	-1.276779313	26.00571356	1
43	-1.363584536	15.55496431	1
44	-3.370201412	17.72935493	1
45	-3.454281415	18.57538832	1
46	-8.85511571	13.91747639	1
47	-10.97907417	22.43697987	1
48	-11.1494087	13.00857535	1
49	-16.14349484	22.2489041	1
50	-16.58290175	15.59992033	1
51	-17.42363869	8.642710899	1
52	-18.14549309	23.14319282	1
53	-20.35881599	23.77317638	1
54	-21.66258174	24.90032386	1
55	-22.03997988	24.20439264	1
56	-23.80257709	29.16352622	1
57	-29.24067454	13.76772805	1
58	-31.15478746	20.68285653	1
59	-32.85708773	25.42954397	1
60	-45.25380886	28.26042186	1
61	-46.44452054	24.41212836	1
62	-48.70906627	19.27218248	0.540924476
63	-56.83841056	14.49275352	0.005404533
64	-59.87236085	18.24795252	0.05816256
65	-261.3934484	48.24976535	3.98E-06

The Brigance Test

The Brigance test is actually several different tests given to children according to their age; a given raw score does not necessarily mean the same thing for a child at fifty months of age as it does for a child at fifty-two months of age. Thus, the first step in employing the Brigance data in an evaluation of pre-k collaborative sites was to convert raw scores to age-normalized composite scores.

The Brigance manual[22] specifies that the appropriate test for a child is determined by that child's rounded chronological age in months, calculated in terms of thirty-day months and rounded up to the next higher month at fifteen days. PEER employed the manual's algorithm to determine each Brigance participant's age at pre- and post-test, and converted raw scores to composite (age-normalized) scores according to tables from the manual.

The Brigance manual's tables do not record exact composite scores for every raw score; some composite scores are simply given as greater or less than some integer value. For this analysis, any composite score given as less than some value X was assigned the value X-0.5; any composite score given as greater than some value X was assigned the value X+0.5. For instance, the composite score ">130" was analyzed as 130.5.

PEER does not have assurance that students were given the correct tests for their Brigance age, or that birthdates were recorded correctly; outliers in the age data suggest that these issues are worth attention.[23] Nonetheless, for this analysis PEER assumed all ages were recorded correctly and all individuals were given the age-appropriate Brigance test.

Because this analysis was designed for the assessment of a population rather than causal generalization from a sample to a population, individuals missing either a pre-test or a post-test were dropped from the analysis; the final analysis was thus based on 1808 pairs of Brigance tests, before and after pre-k, divided among 64 schools. One school had only one individual with both pre- and post-tests, and thus could not be analyzed meaningfully.

Statistical analysis was performed separately on each of the remaining 63 schools, using Pratt's[24] modification of the Wilcoxon signed-rank test.[25] This is a nonparametric test assessing whether two paired distributions are distinct; it effectively asks the probability that a random draw from one ordered distribution outranks a random draw from another. Despite the age-normalization of the Brigance scores, this method was chosen because the Brigance test has not been validated for longitudinal studies of individual students. The null distribution of the test statistic was obtained using exact permutation rather than normal approximation or Monte Carlo sampling; as such, there is no range of uncertainty around the p-values. P-values were Holm-corrected to control for the testing of multiple hypotheses.

Table 2 gives pre- and post-test mean Brigance composite scores for each school, along with the Holm-corrected p-value of the Wilcoxon test for each school. Schools are anonymized and presented in ascending order of p-values.

---

[22] French, B. (2013). Brigance Screens III Technical Manual. North Billerica, MA: Curriculum Associates.
[23] The youngest individual at pre-test was nominally 40 months old (3 years and 4 months); the oldest at post-test was 73 months old (6 years and 1 month). Both these individuals are unusual by reason of age in pre-k.
[24] Pratt, J. W. (1959). Remarks on zeros and ties in the Wilcoxon signed rank procedures. *Journal of the American Statistical Association* 54(287), 655–667.
[25] Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin.* 1 (6): 80–83.

Table 2:

School	Mean pre-test	Mean post-test	Holm p-value
1	98.77419355	121.8225806	3.01E-16
2	87.78378378	96.97297297	1.24E-11
3	90.32	104.4866667	2.30E-10
4	91.25454545	104.8454545	1.44E-09
5	99.37704918	116.4672131	2.02E-08
6	84.14102564	107	1.29E-07
7	99.84090909	107.4488636	1.27E-05
8	100.1428571	112.2285714	1.42E-05
9	86.83333333	99.30952381	1.72E-05
10	88.15	110.9	0.000102997
11	81.78947368	121.0263158	0.000202179
12	88.26315789	97.60526316	0.000499243
13	92.04761905	104.0238095	0.000632286
14	83.40909091	99.63636364	0.000786781
15	86.36842105	101.3157895	0.001308441
16	86.14285714	96.14285714	0.001436949
17	94.32	107.32	0.002644539
18	98.64	103.6066667	0.002883334
19	85.94736842	99.89473684	0.007381439
20	89.33333333	115.9333333	0.008056641
21	90.61538462	108.1538462	0.010498047
22	95.26470588	105.0882353	0.014495278
23	97.66666667	109.5333333	0.015014648
24	97.92857143	106.1428571	0.048828125
25	95.57	103.82	0.048828125
26	100.4	111.45	0.072551727
27	97.57894737	107.5789474	0.072551727
28	96.21428571	104	0.074707031
29	94.91428571	104.1571429	0.080997975
30	96.90789474	102.9736842	0.101742689
31	85.75	98.775	0.101742689
32	96.88461538	104.6923077	0.218093872
33	90.61538462	108.6153846	0.317871094
34	105.1818182	92.18181818	0.380859375
35	86.375	102.625	0.453125
36	102.1176471	109.7941176	0.546447754

37	90.13333333	102.5666667	0.546447754
38	97.16666667	101.1111111	0.546447754
39	89.29220779	92.64285714	0.546447754
40	92.35294118	98	0.546447754
41	94.15384615	106.6153846	0.546447754
42	94.125	101.6875	0.546447754
43	90	96.2	0.8203125
44	104.3636364	101.0909091	1
45	101.6666667	108.9722222	1
46	86.8	98.8	1
47	104.4	109.7	1
48	93.75	119	1
49	110.9	112.925	1
50	88.30555556	87.98333333	1
51	108.5714286	99.57142857	1
52	90.83333333	89.16666667	1
53	91.41509434	91.61320755	1
54	108.1052632	113.0263158	1
55	90.6	90.73333333	1
56	93.68571429	93.31428571	1
57	93.15384615	94.61538462	1
58	92	99	1
59	92.11111111	87.55555556	1
60	93.42307692	101.3076923	1
61	87.5625	94.5625	1
62	88.05263158	88.10526316	1
63	87.55	93.325	1

Classroom Assessment Scoring System Scores

CLASS scores are inherently criterion-referenced and applied to specific classrooms at specific times. As such, very little inference is required to render them useful for evaluative purposes. However, national data[26] indicate that the middle range of the CLASS scale does not actually represent the central tendency of pre-k classrooms, and that the distributions of CLASS scores on individual domains are quite different from one another. Thus, in order to bring this context to the raw CLASS scores, PEER elected to add a comparison to national data.

PEER requested CLASS data for Head Start grantees from the Early Childhood Learning and Knowledge Center (ECLKC), a division of the U.S. Department of Health and Human Services'

[26] E.g., from A National Overview of Grantee CLASS® Scores in 2018. Retrieved October 9, 2019 from https://eclkc.ohs.acf.hhs.gov/data-ongoing-monitoring/article/national-overview-grantee-class-scores-2018. All national data in this section were taken from this site.

Administration for Children and Families. With these data, it would have been possible to precisely compare Mississippi's pre-k collaborative classrooms to national data and to make some statistical comparisons. However, ECLKC staff were unable to provide the requested data in time for the report.

Thus, PEER relied on summary data reported on the ECLKC's website to approximate national percentiles. 2018 summary data – the most recent available from ECLKC – included only means, population standard deviation, minima, and maxima for CLASS dimensions (i.e., the components of a domain score); for the three overall domains, the summary information also included the medians and ninetieth and tenth percentiles.

From this information, it was possible to convert Mississippi's pre-k classrooms' CLASS scores into z-scores. However, because z-scores are not intuitively interpretable to some audiences, PEER approximated percentiles for each Mississippi classroom by assuming that the national data were normally distributed and calculating the cumulative distribution function (i.e., the function giving the probability of an equal or smaller result for each possible value of a CLASS domain) for that normal distribution.

These approximated percentiles are intended for heuristic purposes only; they do not represent empirical 2018 percentiles. Graphics presented on the ECLKC website show that each empirical distribution is approximately normal, though it is impossible to make this visual approximation more precise without access to the underlying data. Table 3 shows that the approximated ninetieth and tenth percentiles are close to reported empirical ninetieth and tenth percentiles, but again, without the underlying data, it is impossible to make a precise comparison of the empirical and approximate distributions.[27]

This method does very slightly underrate any classroom scoring the maximum of 7 in a CLASS domain, because it does not allow for maxima or minima in the scale; however, the underrating runs from slightly less than 0.15% to just over 0.1%. The method would overrate classrooms scoring a 1 in any domain, but no classroom scored so low on any domain.

Table 3:

Domain	Reported tenth percentile	Reported ninetieth percentile	Approximated tenth percentile	Approximated ninetieth percentile
Emotional Support	5.6641	6.45	5.682719015	6.477280985
Classroom Organization	5.2803	6.28	5.300194889	6.299805111
Instructional Support	2.3125	3.71	2.255146639	3.664853361

Table 4 shows the domain scores, z-scores, and approximated percentiles for all Mississippi collaborative pre-k classrooms, rounded to two decimal places; the unrounded data underlie Exhibit 8, page 34. Classrooms are presented without external identification, in order of Emotional Support domain score, in order to protect the students' identities.

[27] It is worth noting, of course, that the empirical distribution obtained in a given year is itself only an approximation of a true underlying distribution, which will by definition be closer to normality than any given year's data insofar as the underlying trait is distributed normally in the population. The fact that a given year's empirical distribution is not precisely normal is not itself a strike against normal approximations.

Table 4:

Class	Emot. Supp.	Class Org.	Instr. Supp.	Emot. Supp. Z	Class Org. Z	Instr. Supp. Z	Emot. Supp. %	Class Org. %	Instr. Supp. %
1	7	6.33	2.75	2.97	1.36	-0.38	99.85	91.29	35.13
2	7	6.58	3.83	2.97	2.00	1.58	99.85	97.72	94.32
3	7	6.08	3.08	2.97	0.72	0.22	99.85	76.36	58.64
4	7	6.42	3.83	2.97	1.59	1.58	99.85	94.41	94.32
5	7	6	3.33	2.97	0.51	0.67	99.85	69.60	74.94
6	7	6.25	2.42	2.97	1.15	-0.98	99.85	87.57	16.31
7	7	6.83	3.58	2.97	2.64	1.13	99.85	99.59	87.02
8	7	6.92	2.83	2.97	2.87	-0.24	99.85	99.80	40.66
9	6.94	7	4.83	2.77	3.08	3.40	99.72	99.90	99.97
10	6.94	6.83	3.17	2.77	2.64	0.38	99.72	99.59	64.87
11	6.94	6.08	2.42	2.77	0.72	-0.98	99.72	76.36	16.31
12	6.94	6.25	2.67	2.77	1.15	-0.53	99.72	87.57	29.90
13	6.94	6.5	3.5	2.77	1.79	0.98	99.72	96.37	83.69
14	6.94	6.83	2.33	2.77	2.64	-1.15	99.72	99.59	12.60
15	6.94	6.17	3.83	2.77	0.95	1.58	99.72	82.86	94.32
16	6.94	6.25	3.83	2.77	1.15	1.58	99.72	87.57	94.32
17	6.94	6.33	4.17	2.77	1.36	2.20	99.72	91.29	98.61
18	6.94	6.25	3.17	2.77	1.15	0.38	99.72	87.57	64.87
19	6.94	7	2.67	2.77	3.08	-0.53	99.72	99.90	29.90
20	6.88	6.08	3.33	2.58	0.72	0.67	99.51	76.36	74.94
21	6.88	6.25	4	2.58	1.15	1.89	99.51	87.57	97.07
22	6.88	6.08	3.67	2.58	0.72	1.29	99.51	76.36	90.16
23	6.88	6.58	4.08	2.58	2.00	2.04	99.51	97.72	97.91
24	6.88	6.75	4.08	2.58	2.44	2.04	99.51	99.26	97.91
25	6.88	6.75	3	2.58	2.44	0.07	99.51	99.26	52.90
26	6.88	6.67	3.25	2.58	2.23	0.53	99.51	98.72	70.10
27	6.81	6	2.42	2.35	0.51	-0.98	99.07	69.60	16.31
28	6.81	6	3.17	2.35	0.51	0.38	99.07	69.60	64.87
29	6.81	6.5	2.83	2.35	1.79	-0.24	99.07	96.37	40.66
30	6.81	6	3.75	2.35	0.51	1.44	99.07	69.60	92.46
31	6.81	6.92	2.83	2.35	2.87	-0.24	99.07	99.80	40.66
32	6.81	6.33	2.5	2.35	1.36	-0.84	99.07	91.29	20.15
33	6.75	6.92	3.5	2.16	2.87	0.98	98.47	99.80	83.69
34	6.75	6.33	2.42	2.16	1.36	-0.98	98.47	91.29	16.31
35	6.75	6.67	2.17	2.16	2.23	-1.44	98.47	98.72	7.54

36	6.75	5.67	3.08	2.16	-0.33	0.22	98.47	36.94	58.64
37	6.75	5.75	3.08	2.16	-0.13	0.22	98.47	44.90	58.64
38	6.75	6	3	2.16	0.51	0.07	98.47	69.60	52.90
39	6.69	6.08	3.33	1.97	0.72	0.67	97.55	76.36	74.94
40	6.69	5.83	3	1.97	0.08	0.07	97.55	53.07	52.90
41	6.69	5.33	2.5	1.97	-1.21	-0.84	97.55	11.41	20.15
42	6.69	5.75	3.25	1.97	-0.13	0.53	97.55	44.90	70.10
43	6.69	5.67	3.17	1.97	-0.33	0.38	97.55	36.94	64.87
44	6.69	6.75	2.58	1.97	2.44	-0.69	97.55	99.26	24.48
45	6.69	5.83	2.75	1.97	0.08	-0.38	97.55	53.07	35.13
46	6.63	6.5	3.33	1.77	1.79	0.67	96.20	96.37	74.94
47	6.63	6.42	4.42	1.77	1.59	2.65	96.20	94.41	99.60
48	6.63	6.5	3.5	1.77	1.79	0.98	96.20	96.37	83.69
49	6.63	6.42	4.5	1.77	1.59	2.80	96.20	94.41	99.74
50	6.63	6.58	2.83	1.77	2.00	-0.24	96.20	97.72	40.66
51	6.63	6.17	3.5	1.77	0.95	0.98	96.20	82.86	83.69
52	6.63	6.5	2.92	1.77	1.79	-0.07	96.20	96.37	47.10
53	6.63	5.83	3.25	1.77	0.08	0.53	96.20	53.07	70.10
54	6.56	6.08	2.83	1.55	0.72	-0.24	93.92	76.36	40.66
55	6.56	5.83	3.75	1.55	0.08	1.44	93.92	53.07	92.46
56	6.56	6.17	4.42	1.55	0.95	2.65	93.92	82.86	99.60
57	6.56	6.92	2.5	1.55	2.87	-0.84	93.92	99.80	20.15
58	6.56	5.75	3	1.55	-0.13	0.07	93.92	44.90	52.90
59	6.56	5.58	2.17	1.55	-0.56	-1.44	93.92	28.63	7.54
60	6.56	5.83	2.25	1.55	0.08	-1.29	93.92	53.07	9.84
61	6.5	6	3	1.35	0.51	0.07	91.23	69.60	52.90
62	6.5	6.08	3.25	1.35	0.72	0.53	91.23	76.36	70.10
63	6.5	6.67	2.5	1.35	2.23	-0.84	91.23	98.72	20.15
64	6.5	6.17	3.5	1.35	0.95	0.98	91.23	82.86	83.69
65	6.44	6.25	2.67	1.16	1.15	-0.53	87.72	87.57	29.90
66	6.44	6	3.5	1.16	0.51	0.98	87.72	69.60	83.69
67	6.44	5.75	2.58	1.16	-0.13	-0.69	87.72	44.90	24.48
68	6.44	6.5	2.67	1.16	1.79	-0.53	87.72	96.37	29.90
69	6.44	6.33	3	1.16	1.36	0.07	87.72	91.29	52.90
70	6.44	6.25	3	1.16	1.15	0.07	87.72	87.57	52.90
71	6.44	6.17	2.83	1.16	0.95	-0.24	87.72	82.86	40.66
72	6.44	5.33	2	1.16	-1.21	-1.75	87.72	11.41	4.05
73	6.38	5.17	2.67	0.97	-1.62	-0.53	83.34	5.31	29.90
74	6.38	6.92	2.67	0.97	2.87	-0.53	83.34	99.80	29.90

75	6.38	5.58	2.83	0.97	-0.56	-0.24	83.34	28.63	40.66
76	6.38	6.25	2.25	0.97	1.15	-1.29	83.34	87.57	9.84
77	6.31	5.42	2.67	0.74	-0.97	-0.53	77.09	16.49	29.90
78	6.31	6.33	3.42	0.74	1.36	0.84	77.09	91.29	79.85
79	6.31	5.75	2.58	0.74	-0.13	-0.69	77.09	44.90	24.48
80	6.25	5.83	2.33	0.55	0.08	-1.15	70.83	53.07	12.60
81	6.25	6.08	3	0.55	0.72	0.07	70.83	76.36	52.90
82	6.25	5.83	2.42	0.55	0.08	-0.98	70.83	53.07	16.31
83	6.25	6	2.33	0.55	0.51	-1.15	70.83	69.60	12.60
84	6.25	4.83	2.5	0.55	-2.49	-0.84	70.83	0.64	20.15
85	6.25	6.17	2.83	0.55	0.95	-0.24	70.83	82.86	40.66
86	6.25	5.92	2.92	0.55	0.31	-0.07	70.83	62.08	47.10
87	6.19	5.5	2.25	0.35	-0.77	-1.29	63.86	22.09	9.84
88	6.19	5.58	3.67	0.35	-0.56	1.29	63.86	28.63	90.16
89	6.19	5.75	3.08	0.35	-0.13	0.22	63.86	44.90	58.64
90	6.19	5.92	3.67	0.35	0.31	1.29	63.86	62.08	90.16
91	6.19	4.75	2.67	0.35	-2.69	-0.53	63.86	0.35	29.90
92	6.13	6.17	3.17	0.16	0.95	0.38	56.41	82.86	64.87
93	6.13	6.17	2.58	0.16	0.95	-0.69	56.41	82.86	24.48
94	6.13	5.42	2.17	0.16	-0.97	-1.44	56.41	16.49	7.54
95	6.13	4.92	2.83	0.16	-2.26	-0.24	56.41	1.20	40.66
96	6.13	5.58	2.75	0.16	-0.56	-0.38	56.41	28.63	35.13
97	6.06	5.67	3.17	-0.06	-0.33	0.38	47.43	36.94	64.87
98	6.06	6.17	4.5	-0.06	0.95	2.80	47.43	82.86	99.74
99	6	5.25	3.08	-0.26	-1.41	0.22	39.82	7.92	58.64
100	6	4	1.5	-0.26	-4.62	-2.65	39.82	0.00	0.40
101	6	4.92	2.67	-0.26	-2.26	-0.53	39.82	1.20	29.90
102	6	6.08	2.5	-0.26	0.72	-0.84	39.82	76.36	20.15
103	6	3.92	3.75	-0.26	-4.82	1.44	39.82	0.00	92.46
104	6	3.83	2.42	-0.26	-5.05	-0.98	39.82	0.00	16.31
105	5.94	5.17	2.17	-0.45	-1.62	-1.44	32.58	5.31	7.54
106	5.94	5	2.42	-0.45	-2.05	-0.98	32.58	2.01	16.31
107	5.94	4.42	2.92	-0.45	-3.54	-0.07	32.58	0.02	47.10
108	5.94	4.58	1.83	-0.45	-3.13	-2.05	32.58	0.09	2.00
109	5.88	5.42	1.67	-0.65	-0.97	-2.35	25.94	16.49	0.95
110	5.81	4.75	2.08	-0.87	-2.69	-1.60	19.19	0.35	5.48
111	5.75	3.92	1.67	-1.06	-4.82	-2.35	14.35	0.00	0.95
112	5.69	4.08	2.17	-1.26	-4.41	-1.44	10.42	0.00	7.54
113	5.69	5	2.08	-1.26	-2.05	-1.60	10.42	2.01	5.48

114	5.69	4.17	2.92	-1.26	-4.18	-0.07	10.42	0.00	47.10
115	5.63	4.5	3	-1.45	-3.33	0.07	7.33	0.04	52.90
116	5.63	5.5	1.67	-1.45	-0.77	-2.35	7.33	22.09	0.95
117	5.63	4.17	2.33	-1.45	-4.18	-1.15	7.33	0.00	12.60
118	5.63	5.42	2.83	-1.45	-0.97	-0.24	7.33	16.49	40.66
119	5.63	4.67	1.75	-1.45	-2.90	-2.20	7.33	0.19	1.39
120	5.5	5.58	2.25	-1.87	-0.56	-1.29	3.07	28.63	9.84
121	5.38	6.17	2.08	-2.26	0.95	-1.60	1.20	82.86	5.48
122	5.38	5.25	2.5	-2.26	-1.41	-0.84	1.20	7.92	20.15
123	5.31	4.83	1.75	-2.48	-2.49	-2.20	0.65	0.64	1.39
124	5.31	5.42	2.08	-2.48	-0.97	-1.60	0.65	16.49	5.48
125	4.94	4.42	1.42	-3.68	-3.54	-2.80	0.01	0.02	0.26
126	4.94	4.67	1.75	-3.68	-2.90	-2.20	0.01	0.19	1.39
127	4.31	3	1.75	-5.71	-7.18	-2.20	0.00	0.00	1.39
128	4.25	2.5	1.67	-5.90	-8.46	-2.35	0.00	0.00	0.95

The CLASS graphic in Exhibit 8, page 34, has a small amount of random displacement on the x and y axes added to point locations for readability, because the relatively small number of possible discrete domain scores, combined with the fact that the x and y axes are transformations of the same data, results in a high amount of point overlap.

MISSISSIPPI DEPARTMENT OF EDUCATION

Carey M. Wright, Ed.D.
State Superintendent of Education

December 11, 2019

Mr. James Barber, Executive Director
Mississippi Joint Committee on Performance Evaluation and Expenditure Review (PEER)
Post Office Box 1204
Jackson, Mississippi 39215-1204

Dear Mr. Barber:

The Mississippi Department of Education has reviewed the draft 2019 *Report on the Early Learning Collaborative Act of 2013: An Evaluation of the Operation and Effectiveness of the Program*. Attached is a response to the draft report, as well as several additional resources that should be informative to PEER's report.

We look forward to seeing the final report released, and we are requesting that our response be included as an addendum to the final PEER report.

Regards,

Carey M. Wright, Ed.D.
State Superintendent of Education

Attachments:
01.Response to PEER Draft Report on Early Learning Collaboratives
02.MDE Research Framework and Agenda
03.ELC Impact Study: Phase I Cross-sectional Analysis
04.Presentation: Phase I ELC Study
05.Poster Presentation: Phase I ELC Study
06.Presentation: Phase II Impact Study on Pre-Kindergarten Programs
07.Preschool Curriculum Consumer Report - 03.25.19
08.Effectiveness Evaluation Summary Report - 02.16.16
09.Head Start CLASS Scores - National Averages

*Attachments to the Mississippi Department of Education's response are available for review in PEER offices at 501 North West Street, Suite 301-A, Jackson, Mississippi between 8 a.m. and 5 p.m.

**In its response, MDE references specific PEER report page numbers. However, the production of the final version of this report caused some page numbers to change. Thus, the page numbers referenced in MDE's response are approximated.

Mississippi Department of Education
Response to PEER Draft Report on Early Learning Collaboratives – November 20, 2019 December 11, 2019

PEER COMMENT	MDE RESPONSE
Page 15. Report references the "McComb Early Learning Alliance".	The correct name is the "McComb Community Collaborative for Early Learning Success".
Page 17. PEER mentions Mississippi meeting nine of the 10 NIEER benchmarks.	Pending the release of the 2018-19 school year report (to be issued in April 2020) from NIEER, Mississippi's ELC program will have met 10 of 10 benchmarks. Also, of the programs around the country, Mississippi's ELC was one of only six states to meet nine benchmarks prior to 2018-19. Only three states met all ten benchmarks prior to 2018-19.
Page 18. PEER recommends that MDE develop a research strategy to demonstrate impact.	MDE has in fact developed a research agenda that can be found at https://www.mdek12.org/OTSS/ORD. This research agenda includes early childhood effectiveness as one of four components. MDE has published and/or presented two short-term impact studies available on the above website. MDE is currently conducting its first long-term impact study evaluating the impact of ELC participation on third grade literacy, mathematics and other outcomes. MDE has attached copies of the first study and a presentation for the second study as part of this response.
Page 19. Report cites Tennessee pre-K study and recommends that MDE conduct a similar study.	MDE notes that the programs included in the Tennessee study were not aligned with the NIEER benchmarks. Tennessee also uses a tiered eligibility regime. Mississippi has a different selection process, which varies by collaborative. Impact results based on these differences with Tennessee would not likely be replicable here in Mississippi.
Page 20. PEER states that the OWL curriculum is neither "evidence-based" nor "research-based".	The United States Department of Health and Human Services' National Center on Early Childhood Development, Teaching and Learning (NCEDTL) published a Curriculum Consumer Report for Preschool in March 2019. This attached report provides review summaries and ratings of comprehensive curricula. Thirteen curricula are reviewed and rated, including evidence of child outcomes. Opening the World of Learning (OWL) earned a Moderate Evidence, or a third

	level out of four levels, that indicate modest child outcomes in one or more areas in quasi-experimental or experimental studies. On page 27 of its 2015 report, "PEER found that students participating in the Gilmore Early Learning Initiative Collaborative and the Clarke County Early Learning Partnership (that both used the OWL curriculum) achieved at least a 498 score (the end-of-the-year target score for exiting prekindergartners) significantly more often than a set of students taught under a curriculum determined by multiple, rigorous controlled studies to have no discernable effect (i. e., the Creative Curriculum). The Gilmore Collaborative students had an adjusted pass rate 21% higher than the baseline group (a significant effect at $p < 0.001$). The Clarke County collaborative students had an adjusted pass rate 20% higher than the baseline group (a significant effect at $p < 0.01$)."

PEER COMMENT	**MDE RESPONSE**
	MDE requires that all collaboratives not only meet the rigor of an evidence- or research-based curricula; but the curricula must also align to Mississippi's comprehensive early learning standards and place a strong emphasis on early literacy. Using the Curriculum Consumer Report, the MDE looked for the curriculum that indicated the highest evidence-based results for child outcomes, and the results proved to be OWL. This curriculum also met the other requirements mentioned above. PEER's objection notwithstanding, OWL is the only curriculum that meets this level of all of the required components that MDE has found. The What Works Clearinghouse has no curricula that meets all the needs mentioned above, and the last time OWL was reviewed was December 2005. The Results First Clearinghouse Database (PEW) has no curricula that meets all the needs mentioned above for the MDE.
Page 21-25. PEER is critical of MDE's method for evaluating effectiveness, stating that it does not adequately measure collaborative or site success.	MDE followed standard benchmarking protocols, working with state and national experts through our Technical Advisory Committee (TAC) to establish effectiveness benchmarks with cut scores common between performance and growth on all instruments/components. MDE has attached a copy of the benchmarking summary report as part of this response.

Page 27. PEER recommends using a randomized control trial in the state.	In the field of Pre-K through grade 12 education, there are very tight controls over research using human subjects. Denying access to desperately needed supports or other vital instruction is never in the best interest of students.
Page 28. PEER states that according to its analysis, 88% of ELC sites demonstrated no impact.	MDE notes that PEER's analysis limited the demographic variables. Using a broader range of demographic characteristics would have strengthened the validity of PEER's findings and may have revealed a positive impact, consistent with the aforementioned MDE studies.
Page 29. PEER found certain sites with significant results should be investigated.	The MDE will investigate sites that have significant results that show unlikely progress with the child assessments versus the CLASS observation results. Sites that have realistic performance on all assessments will be asked to share their strategies with other collaboratives on how they are achieving these results. The site with unusually high pre-assessment KRA scores was noted in the fall of 2018, and MDE conducted the post-assessment because of concerns. This collaborative did not continue in the pre-k program.
Page 32-35. PEER is critical of Mississippi's CLASS performance,	MDE finds that the average of the CLASS scores are as follows: Emotional Support – 6.37, Classroom Organization – 5.79, and Instructional Support – 2.87. The Head Start Early Childhood Learning & Knowledge Center shows

PEER COMMENT	**MDE RESPONSE**
without noting that Mississippi ELCs actually are performing close to national averages.	that the national statistics by domain are as follows: Emotional Support – 6.08, Classroom Organization – 5.8, and Instructional Support – 2.96. Mississippi results are on extremely close or above the national average. MDE is working with ELCs on a regular basis through coaching and professional development to support their staff to increase quality interaction in pre-k classrooms.

Page 40. PEER makes a sweeping statement that it does not have assurance that students were given the correct tests for their Brigance age, or that birthdates were recorded correctly.	MDE assessed 2,040 children with the Brigance assessment. Out of those children, there were three children that were assessed that were not four years old. One child was three years old (only two days from being four years old). Two children turned six years old two days before the end-of-year assessment date.

PEER Committee Staff

James A. Barber, Executive Director

Legal and Reapportionment
Ted Booth, General Counsel
Ben Collins
Barton Norfleet

Administration
Alicia Russell-Gilbert
Deborah Hardy
Gale Taylor

Quality Assurance and Reporting
Richard Boada
Tracy Bobo

Performance Evaluation
Lonnie Edgar, Principal Analyst
David Pray, Principal Analyst
Jennifer Sebren, Principal Analyst
Kim Cummins
Matthew Dry
Samuel Hearn
Matthew Holmes
Taylor Mullins
Sarah Williamson
Julie Winkeljohn
Ray Wright

Performance Accountability
Linda Triplett, Director
Kirby Arinder
Debra Monroe-Lax
Meri Clare Ringer