



MISSISSIPPI DEPARTMENT OF EDUCATION

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

October 16, 2015

James Barber, Executive Director
Joint Legislative PEER Committee
Woolfolk Building
501 North West Street, Suite 301A
Jackson, Mississippi 39201

Dear Mr. Barber,

The Mississippi Department of Education (MDE) has reviewed the PEER Committee report "The Early Learning Collaborative Act of 2013: Evaluation of the Operations and Effectiveness of the Program." This PEER Committee report has not followed the Joint Committee on Standards for Educational Evaluation, which are the national standards for evaluating educational programs.

The MDE has prepared the following responses to each PEER critique.

Section I

PEER Critique: No Minimum Rate of Readiness Set for Continued Funding—"MISS. CODE ANN. Section 37-21-51 (3)(c)(iv)(1972) requires that MDE adopt a minimum rate of readiness on the Kindergarten Readiness Assessment that each prekindergarten provider must meet in order to remain eligible for program funds. As of July 24, 2015, MDE had not adopted a minimum rate of readiness."

MDE Response:

MDE has not yet set a minimum rate of readiness for determining whether providers are eligible for continued funding due to circumstances specific to the Kindergarten Readiness Assessment. These circumstances are explained below.

What concerns MDE more are the motives that PEER ascribes to MDE for this delay, which are not supported by any evidence. In its explanation, PEER states that MDE's failure to establish the minimum rate of readiness prior to program implementation opens the possibility that MDE may set an artificially low rate in order to justify future funding requests to the Legislature. Furthermore, PEER claims that MDE's statement that it will consider the role of growth in calculating the readiness rate is yet another way that MDE will introduce "leniency" in the metric. Both of these critiques are faulty. We can only assume that PEER is operating under at least one of four misperceptions:

1) PEER conflates the readiness benchmark with the minimum rate of readiness. The readiness benchmark represents what children should know and be able to do to be ready for success in Kindergarten. Currently, this benchmark is based on a scale score of 498 on the Mississippi K-3 Assessment Support System, which is primarily a literacy test. A score of 498 at the end of a child's four-year-old pre-Kindergarten year is a rigorous benchmark, as explained in Section II. Had MDE set this benchmark low, this would indeed be cause for concern because then MDE could claim that an artificially high number of children were meeting the mark. Not only did MDE **not** do this, this score was set prior to the first full year of implementation of the collaboratives. More importantly, the readiness benchmark is not the same as the minimum rate of readiness, which refers to the percentage of children in a particular collaborative or provider's program who meet the readiness benchmark at the end of the pre-Kindergarten year.

2) PEER misunderstands the purpose of the minimum rate of readiness. The purpose of the minimum rate of readiness is to hold collaboratives and providers accountable for quality program implementation—not to build a case for expanding the program. The plain language of the text—"minimum rate"—implies that MDE should set this rate at the lowest acceptable percentage of children meeting the readiness benchmark. The crafters of the *Early Learning Collaborative Act* modeled the minimum rate of readiness language off of similar language in Florida's law establishing its Voluntary Pre-Kindergarten program. Because funding is not available to establish an early learning collaborative in every community that seeks one, this language is intended to ensure that precious state dollars are given to the programs that are most effective. If the minimum rate of readiness is set too high, MDE risks eliminating effective programs; if the minimum rate of readiness is too low, MDE risks missing the opportunity to fund more effective programs. To try to use the success of funded collaboratives on this measure as an indicator of the effectiveness of the legislation as PEER's critique states would indeed create strange incentives. Such a usage is neither expressed nor implied anywhere in the legislation.

3) PEER lacks a basic understanding of the role of baseline data in setting accountability metrics. PEER's suggestion that MDE could have set a defensible minimum rate of readiness prior to establishing baseline data for a new assessment belies a lack of knowledge about the responsible creation of accountability systems. The

process of setting a minimum rate of readiness is a complicated one that required MDE to first establish a measurement of school readiness. Prior to 2013, Mississippi had neither a school readiness assessment nor even a statewide definition of school readiness for any of the five accepted domains of early childhood learning.

In 2013, the Mississippi Legislature passed both the *Early Learning Collaborative Act* (ELCA) and the *Literacy-Based Promotion Act* (LBPA). The LBPA required MDE to implement a statewide K-3 literacy assessment for the purpose of determining whether K-3 students were on track for meeting the state's reading benchmark by the end of third grade. This K-3 literacy assessment had to incorporate a Kindergarten readiness assessment for literacy to produce baseline data for each student.

The ELCA required MDE to adopt a minimum rate of readiness for Kindergarten if MDE adopts a "statewide Kindergarten screening that assesses the readiness of each student for Kindergarten." It does not require that this assessment be the same as the assessment implemented under the LBPA. However, due to both time and resource constraints, MDE issued a single RFP for a K-3 literacy assessment that would also be robust enough to serve as a Kindergarten readiness assessment. The selected assessment is the STAR Early Literacy assessment, now called the Mississippi K-3 Assessment Support System (MKAS²).

In the absence of an adopted school readiness definition and within the context of having only the MKAS² as a statewide Kindergarten entry assessment, MDE chose the only measurable benchmark available as a school readiness indicator—a score on the MKAS² that was supported by research to indicate a child's chance of school success. A committee of practitioners chose a scale score of 498 as the readiness benchmark in summer 2014, prior to the first full year of implementation for collaboratives. (More on this decision in Section II.)

After MDE selected a school readiness benchmark, it could then determine what percentage of ready children each collaborative and each provider must meet in order to receive continued funding. Setting accountability requirements is necessarily a matter of values, research, and judgment as accountability metrics are only meaningful in the context in which they are designed and used. Accepted practice in education is to establish a baseline of scores on any assessment used for accountability purposes prior to setting accountability requirements, such as the high-stakes minimum rate of readiness referenced in the ELCA.

MKAS² was first piloted in the spring of the 2013-14 school year for K-3 students. Because so few collaboratives were operational at this point in time (the collaboratives were selected in December 2013), very few collaborative classrooms (only 6 of the 111 classrooms funded) participated in this pilot; participation was the result of circumstance rather than any sampling design that could yield meaningful baseline test score data. Consequently, the field test among collaborative classrooms was not sufficient to set a baseline of expected performance for programs. Instead, the field test

enabled MDE and the collaboratives to understand the process needed to test every child on a computer-based exam when not every child was in a facility with the technology to take the exam.

The first year in which a sufficient number of collaborative classrooms administered a fall and spring readiness assessment was 2014-15. As a result, MDE is relying on best practice in education to use assessment results from 2014-15 to set the minimum rate of readiness on the MKAS². A standards validation committee underwent a process to support the benchmark score of 498 on the Star Early Literacy Assessment as the desired score for students exiting Pre-K. A score of 498 at the end of Pre-K means that the student has mastered 70% of the reading indicator skills needed and supports that the student is on track to meet the end-of-grade 3 reading proficiency expectations. This Kindergarten Readiness Assessment data is used in the beginning of the year to inform teachers to implement interventions and strategies to help children to strengthen literacy and a few numeracy skills. The data at the end of the year is used to determine their growth over the year in literacy and numeracy skills.

This work is expected to conclude within the next year. By that time, MDE also hopes to have adopted a statewide definition of school readiness as well as multiple, aligned measures that can assess each enrolled child's readiness. MDE expects to incorporate the finalized minimum rate of readiness for funding decisions beginning in the 2016-17 school year.

4) PEER does not understand the importance of growth in measuring educational effectiveness. A central assertion of the report authors is that a growth-based metric confounds measuring a program's effectiveness. A well-designed and fair system should include multiple indicators, including both status and growth/improvement. These designs assist in determining the degree to which the program is being effective in both specific levels of performance, and improvement both to and beyond the targeted level. Furthermore, to the extent that equity (or improving performance of at-risk students) is a priority outcome, indicators must be included that signal such progress and reward ongoing efforts. To this point, the Council of Chief State School Officers (CCSSO) document entitled *CCSSO Roadmap for Next-Generation State Accountability Principles (CCSSO, 2011)*, developed by a representative group of state education leaders and experts in the field of assessment and accountability, explicitly promotes multi-indicator designs.

Section II

PEER Critique: MDE awarded funding to four collaboratives that utilized a prekindergarten curriculum found through rigorous research to have “no discernable effects” on student learning.

MDE Response:

Each collaborative selected a researched-based curriculum and assessment based on the children's needs in the community, that had a literacy focus, and was aligned to all areas of the Early Learning Standards. All collaboratives met the Early Learning Collaborative Act of 2013 requirements in their curricula and assessment selections. With 2014-15 being the first full year of ELCA implementation, administrators and teachers have been working toward student mastery of the Early Learning Standards through the selected curriculum.

The National Association of the Education of Young Children describes indicators of effective curriculum likely to promote positive outcomes for all young children. These indicators are: children are active and engaged; goals are clear and shared by all; curriculum is evidenced based; valued content is learned through investigation, play, and focused, intentional teaching; curriculum builds on prior learning and experiences; curriculum is comprehensive; professional standards validate the curriculum's subject-matter content; and the curriculum is likely to benefit children (NAEYC & NAECS/SDE (2003)).

PEER referenced the absence of Creative Curriculum in the Results First Clearinghouse Database. OWL, which is identified in the PEER critique as "the curriculum used by classes where the students performed better," was also absent in this Database. Collaboratives met the Request for Proposal's criteria for selection of both of the curricula.

Gilmore Early Learning Collaborative and Clarke County Early Learning Collaborative use the OWL Curriculum as well as Hatch, which is a technology-based program focused on literacy and numeracy. Hatch is used in the classroom on a daily basis and evaluates the students through computer games and creates reports for the teacher on the accomplishments of the students and their learning. Because of this natural relationship of students using the Hatch program, they could be more accustomed to using a computer. Therefore, the students could be more prepared to succeed on a computer-based assessment than those students who do not use Hatch on a daily basis.

Section III

PEER Critique: PEER found that pre-Kindergarteners in the collaboratives achieved the end-of-the-year target score or higher on the Kindergarten Readiness Assessment less than those enrolled in other public pre-Kindergarten programs. Students participating in two collaboratives that used the Owl curriculum achieved the end-of-the-year target score or higher on the Kindergarten Readiness Assessment significantly more often than students taught under the Creative Curriculum (which was determined by multiple, rigorous controlled studies to have no discernable effect on student learning).

MDE Response:

The range of target scores takes into account the variation of the collaboratives with different sites as each collaborative's composition of classrooms includes school district, and/or Head Start, and/or a childcare center. Childcare and Head Start centers have migrated to following testing protocols of school districts. For most collaborative classrooms, curriculum and assessment selections prior to fall 2014 were not governed by MDE. As collaboratives began serving children in fall 2014, this was the first interaction of coordination between partners following grant requirements (curriculum sharing, professional learning communities, conducting screeners and assessments, resource sharing, etc.). All collaboratives had previous agreements working towards an early childhood focused community plan to increase awareness of early childhood education. Many administrators and educators in the collaboratives spent the first portion of the 2014-15 school year blending curricula and formative assessments in instruction as well as ensuring the following of their licensing agency (e.g., childcare, Head Start, MDE) guidelines AND the ELCA.

Title I Pre-K classrooms have been established for many years in districts and for the most part have been in the practice of blending curricula and assessment in instruction for quite some time, ensuring the following of the MDE guidelines only. Collaborative partners (including childcare and Head Start centers) are administering MKAS² testing and screening tools for the first time as a collective group. The 2014-15 school year was the first year the MKAS² was formally administered. Through this administration, there will be data points that will help form literacy instruction in the classroom for pre-K and Kindergarten classrooms. Keep in mind that these scores are based on the performance of a four-year-old child in a 20-minute window at the end of the school year. The assessment scores were designed in the Early Learning Act language as a way to assess readiness for Kindergarten. After the summative assessment in the fall, teachers use the information to influence the intervention provided through activities in the classroom to strengthen skills.

The expected end-of-year scale score for four-year-old students is 498. This scale score at the end of Pre-K means that the student has mastered 70 percent of the targeted early literacy skills needed and supports that the student is on track to meet the end of grade 3 reading proficiency expectations. A scale score value of 498 falls within the Late Emergent Reader Classification Level. The 2014-15 Kindergarten Readiness Assessment results can be used to determine if each student has met the expected performance target of 498 scale scored points. This is a proxy for "basic" mastery of early literacy and should be seen as the minimal score needed to enter Kindergarten.

The scores that were set as the target score for pre-K and Kindergarten were identified as a score on the projection of growth where the child would be a successful reader by the third grade. The score indicated for the third grade target is 926 and is in the progression line with the percentages set for pre-K and K target scores.

Section IV

Technical Issues

MDE Response:

The method outlined in page 2 of the PEER report immediately suggests the report's title has no relationship with the actual method used to draw those conclusions reported on pages 29-31 of the report. The PEER "evaluators" did not conduct a mixed-method design (Tashakkori & Teddlie, 2003) necessary to collect independent, replicable data on the operations of the collaborative. The Joint Committee on Standards for Educational Evaluation (JCSEE) (Yarbrough, et al., 2011) clearly states within the Accuracy Standards, "the procedures of the evaluations should be monitored and described in enough detail, so that they can be identified and assessed." No information was provided on how the evaluators planned, conducted, and analyzed data associated with the implementation of different early learning curriculum. The evaluators conducted a desktop review of literature in lieu of designing a scientifically-based approach sufficient to support their comparative analysis.

The general finding that variance in performance outcomes was observed within the collaboratives is supported by the outcome data. The Department of Education, using unattenuated scale scores, observed a range of performance when the unit of analysis was fixed at the collaborative level. This variance suggests, in the aggregate, that eight (72.7%) of the eleven grant recipients met or exceeded the benchmark established by the MDE. Thus, the statement in the conclusion on page 30 can be corroborated in the outcome evidence. Unfortunately, as the PEER conclusions clearly violate JCSEE standards U1 Evaluator Credibility, U6 Meaningful Processes and Products, P4 Clarity and Fairness, A1 Justified Conclusions and Decisions, A6 Sound Designs and Analysis, A8 Communication and Reporting, and EA2 Internal Metaevaluation, any inferences about the Early Learning Collaborative effectiveness articulated in this report should be disregarded.

The below table enumerates the aforementioned statements:

JCSEE Standard	Technical Violation
U1 Evaluator Credibility	The evaluators fail to have the professional qualifications in the area of educational program evaluation, publications in scholarly journals associated with evaluating early childhood programs, utilization of subject matter experts in early childhood, and evaluating the psychometric characteristics of computer-adapted testing necessary to meet this standard.

JCSEE Standard	Technical Violation
U6 Meaningful Processes and Products	The evaluators failed to utilize socio-economic data of the actual students but rather relied on proxy data. No data was collected on the operational fidelity of the curricula implement by the subgrantees. No information was analyzed on the psychometrics of the assessment, the key dependent variable cited in the report. No comprehensive literature review demonstrating the scope and selection of the evaluation is supported within the larger body of research conducted throughout the world on early childhood programs. No theoretical framework or theory of action was used to support the evaluation’s design. No data assumptions and limitations were comprehensively identified by the evaluators. No technical testing was conducted to support the analytical model used by the evaluators.
P4 Clarity and Fairness	The evaluators’ scope is too narrow to support the conclusions; alternative explanations to the cited differences in performances are not stated.
A1 Justified Conclusions and Decisions	The evaluators’ conclusion about the comparative group is unsupported due to failure by the evaluators to ensure equivalency of known exogenous factors impacting measures of latent constructs (achievement); conclusions associated with curriculum effectiveness are confounded by construct irrelevant characteristics both omitted and exacerbated by the use of proxy data that does not fit the actual dataset. The evaluators’ use of a “four year longitudinal study” reported by the MDE is inaccurately cited.
A6 Sound Designs and Analysis	The evaluators’ dismissal of the assessment’s psychometric evidence is not based upon documented research; the evaluators’ attenuation of interval level data distorts the results on effectiveness; the method used to develop a “pseudo” control group for comparative purposes did not establish equivalence of the groups on key independent factors; the evaluators failed to examine the dependent variable, which supports the use of more sophisticated and robust parametric analytics [reducing Type 1 errors]; the evaluators failed to use econometrics in making return-on-investment (ROI) statements.
A8 Communication and Reporting	The evaluators’ purpose fails to align with the conclusions; the evaluators failed to enumerate the research question and subordinate questions prior to the beginning of the activity; the evaluators refused to accept MDE assistance in reviewing the methodology prior to the evaluation; the evaluators refused to disclose the statistical software planned for use during the evaluation; the evaluators did not identify the SMEs to be utilized; the evaluators did not request permission to conduct qualitative reviews; the evaluators’ representative made no claim of an intent to conduct a fair, impartial evaluation.

JCSEE Standard	Technical Violation
EA2 Internal Metaevaluation	The evaluators failed to conduct sufficient review of the literature to support their efforts and findings.

References

NAEYC & NAECS/SDE (2003). *Early childhood curriculum, assessment, and program evaluation: Building an effective, accountable system in programs for children birth through age 8. Joint position statement*. Retrieved from www.naeyc.org/cape.

Tashakkori, A. & Teddlie, C. (2003). *Handbook of Mixed Methods in Social & Behavioral Research*. Thousand Oaks: Sage.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.

Conclusion

This PEER report is the second PEER Committee report produced in 2015 that has not followed the Joint Committee on Standards for Educational Evaluation, which are the national standards for evaluating educational programs. Therefore, any conclusions or recommendations contained in this PEER report lack merit.

Sincerely,



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