



Achieving Academic Equity and Excellence for Black Boys (AAEEBB)

Year Two Evaluation Report

Office of Research, Planning, and Program Evaluation

March 2024

MARYLAND STATE DEPARTMENT OF EDUCATION**Carey M. Wright, Ed.D.**

State Superintendent of Schools

Geoff Sanderson

Deputy State Superintendent
Office of Accountability

Dr. Deann Collins

Deputy State Superintendent
Office of Teaching and Learning

Wes Moore

Governor

MARYLAND STATE BOARD OF EDUCATION**Clarence C. Crawford**

President, Maryland State Board of Education

Joshua L. Michael, Ph.D. (Vice President)

Chuen-Chin Bianca Chang, MSN, PNP, RN-BC

Susan J. Getty, Ed.D.

Dr. Monica Goldson

Nick Greer

Dr. Irma E. Johnson

Dr. Joan Mele-McCarthy, D.A., CCC-SLP

Rachel L. McCusker

Samir Paul, Esq.

Abhiram Gaddam (Student Member)

Table of Contents

Reflections from the Advisory Council	4
Executive Summary	5
Introduction	8
Evaluation Goals and Framework	11
Evaluation Plan and Design	13
Process Evaluation Findings	16
Impact Evaluation Findings	29
Analysis of Program Budgets and Expenditures	36
Limitations	38
Conclusion	39
Recommendations	40
Appendix	44

Reflections from the Advisory Council

As we reflect on two years of the Achieving Academic Equity and Excellence for Black Boys (AAEEBB) pilot, I'm encouraged by our progress. The Maryland State Department of Education has expressed a long-term commitment to this effort which I'm certain will continue to show movement towards meeting the objectives. If our state is serious about transforming school culture for Black boys in ways that improve academic performance and decrease disproportionality in discipline, the responsive effort must have the resolve necessary to counter the issue. When the vision was laid for this work in 2020, it was anchored in an understanding that there would be multiple and consistent supports for the initiative. As can be seen in the data, the efforts show promise, however consistency, or the lack thereof, can impact the mission.

Moving into year three, one can see the types of programming that address the challenge. A focus on quality instruction and engagement via culturally responsive and equity centered practices is a meaningful endeavor. It is clear that our professional learning has had an impact. Additionally, the pro-social mentoring, leadership, and social emotional learning experiences for our boys is showing growth. The programming increases the students' sense of personal and academic self as well as demonstrates the importance of trusted relationships and high expectations.

We know this work will continue to take time and will be challenging to measure, but we are observing the seedlings of our efforts. It is important for us to remember that the schools will follow and replicate the state's lead, will, and commitment. I hold optimism that the Maryland State Department of Education will continue to explore this issue and research and seek the solutions necessary to change the trajectory of the Black male student experience. We must remain steadfast in our efforts. In the words of former State Board member and Achieving Academic Equity and Excellence for Black Boys lead organizer, Dr Vermelle Greene, "If our Black boys are in trouble, so are we all."

Daryl C. Howard, Ph.D.

AAEEBB Advisory Council Chairperson

Executive Summary

The *Achieving Academic Equity and Excellence for Black Boys* (AAEEBB) grant is designed to address the academic and socioemotional needs of Black boys in Maryland Public Schools. In year two of the grant (2022-2023), 14 schools were each awarded \$50,000 to implement programs that addressed recommendations from a state-level task force made up of members of the State Board of Education, personnel from the Maryland State Department of Education (MSDE), local education agency (LEA) administrators, and others. This evaluation report analyzes program implementation and impact for the 14 AAEEBB schools in year two of grant funding.

SELECTION OF STUDENTS AND SCHOOLS

State administrative data show that, consistent with the aims of the Task Force, AAEEBB programs are reaching Black boys who face academic and socioemotional challenges in Maryland public schools.

- Consistent with data presented in the Task Force report, in 2021-2022 the 14 schools had significantly higher suspension rates for Black boys (14%) compared to the average for Maryland Public Schools (6%). AAEEBB schools also had significantly lower perceptions of school climate among both students and educators.^{1 2}
- Within the 14 schools, students participating in AAEEBB programs were more likely to be economically disadvantaged (54%) compared to other Black boys (41%) and other students (37%) in program grades. Participating students also had lower state standardized assessment scores, equal or higher rates of course failure, and higher rates of suspension.

PROGRAM IMPLEMENTATION

AAEEBB schools adopted a variety of Task Force recommendations, but most implemented some combination of mentorship programs (13 of 14 schools) along with professional learning and development focused on conflict resolution, anti-bias and/or culturally responsive teaching practices (11 of 14 schools).³ Data submitted by schools suggest that in year two AAEEBB programs were implemented in a way that is consistent with what is known about how to tailor effective programming to support the academic and socioemotional needs of Black boys, though many schools faced implementation challenges that were perceived to have affected their ability to provide the strongest programming.

¹ Report available at:

<https://marylandpublicschools.org/stateboard/Documents/2021/0427/MSDETransformCultureforBlackBoy.pdf>

² The Maryland Climate Survey includes questionnaire modules intended to measure perceptions of the school environment, including safety, the behavior of students and relationships among teachers and students. It is a core part of the state's accountability system.

³ Because schools were asked to implement specific Task Force recommendations, this report focuses on mentorship programs implemented in response to recommendations 1.3 and/or 1.4, professional learning and development programs implemented in response to recommendations 1.1, 2.4, and/or 3.1, and curricular/academic support implemented in response to recommendations 3.2, 3.3, and/or 3.6. See Appendix and/or Task Force report for more information on recommendations.

- Across all AAEEBB schools, mentorship programs had a ratio of 1.8 mentees to mentors, and 9 of the 13 schools implementing mentorship programs had a ratio of 3.0 or less. In nine schools, mentorship sessions were held at least weekly.
- Eight of the 11 schools implementing professional learning and development (PLD) sessions did so once a month or more frequently, and roughly half (six) required staff participation at most sessions offered.
- 12 of 13 schools selected mentors based on prior knowledge/familiarity with students, and 3 schools required a formal application process for mentors. About half (six) of schools experienced challenges recruiting qualified mentors, and more than half (eight) provided training to mentors. Mentor recruitment, selection, and maintaining a consistent mentor pool were singled out as sources of challenge by many schools in year two. Several schools described efforts to improve the “match” between mentors and mentees, such as requiring applications and statements of interest.
- Seven of 11 schools implementing PLD sessions on conflict resolution, anti-bias and/or culturally responsive teaching practices provided participants with opportunities to apply lessons in practice.

Similar programs such as the African American Male Achievement Initiative in Oakland and the Expanded Success Initiative in New York City include core features such as sustained and consistent opportunities for participant engagement, rigorous recruitment of mentors and opportunities for participants to apply professional learning and development in practice. Indeed, these features of educational programs have shown to be associated with improved outcomes among students and other participants, and in this way AAEEBB programs have shown to be aligned with the evidence base.⁴

PROGRAM IMPACT

To better understand program impact, the evaluation team used state administrative data and a quasi-experimental research design to examine trends in academic, attendance, and behavioral outcomes for AAEEBB students from 2021-2022 to 2022-2023. The impact analysis found that in year two, AAEEBB students reduced their rate of absenteeism, although their academic outcomes were mixed.

Specifically:

- The quasi-experimental analysis, which represents the strongest evidence of the impact of AAEEBB programs, shows the rate of absenteeism for mentorship students was reduced by 4 percentage points, which is a considerable improvement from the absence rate of 8% in 2021-2022. At the same time, there was a statistically significant reduction (0.13 standard deviations) in mentorship students' ELA assessment scores. As a hypothetical illustration of this magnitude, it is similar to the typical third grader in 2022 seeing their scale score drop from the 50th percentile of students to the 46th percentile.

⁴ See appendix for more information on research.

- To understand whether any specific aspect of schools' AAEEBB programs was associated with improved outcomes for students, the evaluation team tested four separate hypotheses using a descriptive (i.e., not quasi-experimental) outcome trends analysis. The four hypotheses are related to aspects of program implementation that have shown to be associated with improved outcomes in previous studies. Although only suggestive, the analysis in this report does not find consistent evidence in favor of any of the four hypotheses tested.

Introduction

One of the most enduring educational challenges in the United States is the disparity in outcomes for Black boys. On the one hand, data consistently show that Black boys have lower proficiency rates on state standardized assessments, as well as lower graduation rates, compared to other student groups (MSDE, n.d.) At the same time, there is growing recognition that Black boys face unique barriers that manifest in daily schooling experiences. Between 2015 and 2018, Black students were suspended or expelled at two to three times the rate of other student race/ethnicity groups, with boys at higher rates than girls (MSDE, n.d.) Research has documented that educational experiences for Black boys are strongly predictive of worse education and employment prospects later in life (Johnson and Jackson, 2019; Bacher-Hicks, et al., 2019). In order to ensure students from all backgrounds have equitable access to safe and nurturing educational opportunities, high-level efforts have long been aimed at addressing inequities for Black boys (MSDE, n.d.)

In the summer of 2020, the Maryland State Board of Education (MSBE) convened the Task Force on Achieving Academic Equity and Excellence for Black Boys (AAEEBB) to explore these inequities and to develop evidence-based strategies for improving educational experiences and outcomes for Black boys in Maryland. The Task Force was composed of 22 members drawn from the State Board of Education, staff from the Maryland State Department of Education, teachers and administrators from Maryland Public Schools, and researchers. Meeting monthly from July 2020 to March 2021, the Task Force formed study groups organized around the themes of social and emotional behavioral support, recruiting and training skilled, competent teachers and administrators, and curricula and instruction. In the summer of 2021, the Task Force issued 16 recommendations for schools and districts to tailor educational programs for Black boys. The recommendations were provided with a call to action and, beginning in the 2021-2022 academic year, 14 schools chosen by local superintendents who expressed interest in the program were given grant funding to adopt two or more recommendations issued by the Task Force.

This report summarizes the experiences of AAEEBB programs during the second year of program implementation (SY 2022-2023). The year one report documented key information on schools' programs, including the Task Force recommendations they chose, program features, and goals and monitoring efforts.⁵ This evaluation report documents evidence on the impact of schools' programs, as well as for process of program implementation.

PROGRAM DESCRIPTION

The aim of the AAEEBB Task Force was to issue recommendations that schools and districts could use to tailor programming to meet the academic and socioemotional needs and behaviors of Black boys. Three study groups were formed to issue several recommendations within broad areas; the first was centered around social, emotional, and behavioral supports, the second was dedicated to the topic of recruiting and training skilled, competent teachers and administrators, and the third was for curricula and instruction. The full list of recommendations can be found in Appendix Table 2. At the end of year one, MSDE's Office of Research, Planning, and Program Evaluation issued a report that summarized program activities. It was found that almost every school had implemented a mentoring program, which typically involved students meeting regularly one to one or as a small group. Sessions could encompass a range of activities, including homework support, tutoring on soft skills and collective

⁵ Report available at: <https://marylandpublicschools.org/about/Documents/ORSDU/AAEEBBPilotReport2021-2022.pdf>

reflection on challenges and conflicts faced in the school and/or home environment. Many tutoring programs included teambuilding and/or group activities, such as field trips or athletics. Most schools also implemented some form of professional learning and development program for staff on anti-bias and conflict de-escalation, as well as culturally responsive teaching and learning strategies. These programs could take the form of structured professional development programs but could also include a book study and/or speaker series.

In the 2022-2023 academic year, the same 14 pilot schools continued implementing their AAEEBB programs. The average size of the grant per school was \$50,000 for the year. Schools were responsible for implementing Task Force recommendations subject to constraints of time, staffing, and budget. As shown in Table 1, schools chose a range of recommendations to adopt. Following patterns from the first year, 13 of 14 schools implemented mentorship and/or Rites of Passage programs in response to Task Force recommendations 1.3 and 1.4. Additionally, 11 schools implemented some form of professional learning/development program for staff. These programs could have taken the form of conflict de-escalation training (recommendation 1.1, three schools implemented programs), training on culturally responsive teaching practices (recommendation 2.4, two schools implemented programs), or professional learning to support culturally relevant pedagogy/teaching and/or anti-bias classroom practices (recommendation 3.1, eight schools implemented programs). Interestingly, none of the 14 schools adopted recommendations related to recruiting, hiring, retaining, or otherwise promoting equity in the teaching staff. On average, schools implemented programs to address two to three recommendations, with one school implementing one recommendation and one school implementing four recommendations.

Table 1 – Number of Schools by AAEEBB Task Force Recommendation

Rec. Number	Recommendation Description	N
1.1	Require de-escalation and other evidence-based intervention and training for all staff.	3
1.3	Coordinate structured mentoring programs tailored to meet the social/emotional needs of Black boys in grades K-12.	13
1.4	Implement a Rites of Passage program for Black boys in grades 9-12.	1
2.4	Provide all teachers and school-based administrators, along with all other district personnel, continued professional development on culturally responsive teaching practices and methods to support the academic, social-emotional, and developmental needs of Black boys and young men.	2
3.1	Provide professional learning to support the implementation of Culturally Relevant Pedagogy/Culturally Responsive Teaching and Anti-Bias Practices in the classroom.	8
3.2	Address ongoing achievement gaps by using the science of reading (systematic phonics instruction, explicit instruction in phonemic awareness, methods to improve fluency, and ways to increase comprehension).	1
3.3	Address ongoing gaps in math and continual math decline through advancing grades by using standards-based, real-world math instruction infused with science, technology, engineering, and math (STEM) activities.	2
3.6	Districts encourage individual co-ed schools to create single gender classes across grades or for selected subjects. Districts may also consider approving charter schools with single-gender classes for boys and girls	2

Note: Information on recommendations was submitted as part of schools' end of year two survey. Schools could choose multiple recommendations and were encouraged to select two. The full set of Task Force recommendations are listed in Appendix Table 2.

Evaluation Goals and Framework

This section describes the overarching goals for the year two evaluation, and how goals informed the development of a framework for understanding the implementation and impact of AAEEBB programs. Broadly, goals for the evaluation are to describe schools' AAEEBB programs, to understand whether programs improved student- and/or school-level outcomes, and to provide evidence for why programs may or may not have improved outcomes. It is important to note this process necessarily involves making generalizations across the 14 AAEEBB schools; for example, AAEEBB schools offered professional training and development opportunities of very different types and with different aims. Therefore, the evaluation framework must be able to account for the intended aims and strategies of a diversity of programs across schools. This section describes this process in detail.

IDENTIFYING OUTCOMES OF AAEEBB PROGRAMS

AAEEBB programs can be understood to have two broad and interrelated goals: 1) improving academic, behavioral and socioemotional outcomes for Black boys, and 2) improving school climate, as well as the curricular, instructional, and administrative barriers/facilitators to improved outcomes for Black boys. This section outlines outcome measures chosen to evaluate the AAEEBB program, and Appendix C gives additional details about how outcome measures are calculated using existing data sources.

The evaluation uses the following as outcome measures at the student level for the impact evaluation:

- Rates of course failure and state assessment scores as measures of academic achievement.
- Suspension and absence rates, which are broadly intended to capture behavior and socioemotional competencies. While undoubtedly, and as noted by the AAEEBB Task Force report, behavior and school attendance are often influenced by a large number of schooling and non-schooling factors, they have also shown to be highly correlated with socioemotional measures such as persistence, grit, and self-regulation (Jackson, 2018; West et al., 2016). Given the difficulties of measuring socioemotional outcomes using readily available data sources, for the purposes of this evaluation they are considered to be outcomes of interest.

Importantly, and as noted by the Task Force, because individual outcomes are also a reflection of systemic and environmental barriers/facilitators, the student-level analysis is combined with a school-level investigation. School-level measures include:

- School climate survey scores for students and educators. School climate was measured in 2022 and 2023 for all Maryland Public Schools, using separate student and educator surveys to capture perceptions of school community, the physical and behavioral environment of the school, relationships among staff and students, and school safety.
- Racial disproportionality in suspension rates. The extent to which Black boys are suspended compared to boys of other races.
- Racial composition of teaching staff. The proportion of teachers that are Black at the school level, compared to the proportion of the teaching staff that are other races.

AAEEBB PROGRAM IMPLEMENTATION

This evaluation sought to understand not just whether AAEEBB programs improved outcomes, but also to generate evidence for why schools were able to improve outcomes; in other words, to understand whether specific aspects of schools' implementation of AAEEBB programs would be associated with

improved outcomes. To understand aspects of implementation that may predict better outcomes for students, the evaluation team conducted a review of the research literature on the types of programs that AAEEBB schools were known to be implementing. Appendix B provides a summary of the literature.

Using the literature review as a basis, the evaluation team selected factors such as the frequency and intensity of schools' AAEEBB programs, as well as specific program features such as how schools recruited mentors and whether professional learning and development programs provided opportunities to apply lessons learned in practice, that might predict differential outcomes for students and schools. Because the evidence base was relatively limited, it was decided that additional descriptive information such as the context and timing of AAEEBB programs could be informative.

Evaluation Plan and Design

This section describes the evaluation plan and design, which consists of two parts – process and impact. Understanding whether programs worked as intended is best achieved through an impact evaluation, while describing and understanding why programs may or may not have worked is best achieved through a process evaluation. This section details each of these parts of the evaluation plan and design.

PROCESS EVALUATION

There were two specific challenges to deciding on measures of implementation at the outset. For one, implementation data would need to be able to be collected directly from schools rather than from a central repository. Secondly, to be most informative, implementation data would need to be generalizable enough to learn about differences among AAEEBB programs collectively.

While the choice of outcome measures is designed to yield evidence for whether schools' AAEEBB programs improved outcomes in year two, an approach was needed for better understanding why programs might or might not improve outcomes. Building on the year one report in which an open-ended survey yielded rich description of schools' AAEEBB programs, the year two survey included a series of multiple-choice questions designed to yield information that could be used to test hypotheses related to the effectiveness of different program components. These questions were designed to yield information on the frequency/intensity and types of engagement for participants in schools' AAEEBB programs, and to better understand qualitative aspects of programs, such as the factors that led to the recruitment and selection of mentors and whether professional learning and development programs provided opportunities to apply lessons in practice.

The AAEEBB school survey was administered from June to July of 2023.⁶ Key information from the AAEEBB school survey used in this report includes:

- The grade levels of students participating in schools' AAEEBB programs.
- Information on characteristics of schools' AAEEBB programs relevant for understanding how programs were implemented.
- Results of goals/outcomes associated with school monitoring.
- Successes and challenges in implementing programs.

IMPACT EVALUATION

To measure program impact, the evaluation team used student- and school-level data from MSDE's multi-year data warehouse, for the 2021-2022 and 2022-2023 school years. This section describes the process followed to group student- and school-level data for the impact evaluation. Once groups were identified, the dataset was constructed and a research design was developed, which this section also details.

⁶ Using the year one report and schools' year two implementation plans submitted in the fall, a questionnaire was first drafted and shared internally, and feedback led to some preliminary modifications. The questionnaire was then pilot tested with all members of the wider Research, Planning, and Program Evaluation team. This led to further modifications and changes, and the final draft was administered starting in June of 2023.

Grouping Student-Level Data

As stated, findings from the year one report indicated schools implemented a mix of targeted programming for students (such as mentorship) and PLD for teachers and other staff. Evaluating the impact of the former requires knowing the students who participated in targeted programming. Evaluating the impact of the latter was expected to be more challenging, for several reasons. For one, the staffing data collected by MSDE has limited measures on knowledge, attitudes, and/or skills, making it difficult to establish the direct impact of PLD programs. Second, PLD programs may have been offered for a large number of staff in a mix of voluntary and mandatory formats throughout the year, making it difficult to obtain a precise list of participants. Third, even if PLD programs were targeted toward specific staff members, they would be expected in turn to interact with students throughout the school and not just those in AAEEBB programs. For these reasons, the impact evaluation focuses primarily on measuring outcomes for students participating in mentorship/Rites of Passage programs implemented in response to recommendations 1.3 and 1.4.

An advantage of choosing mentorship students is that there is a high degree of certainty that their outcomes should be affected by the program. A disadvantage is that there are other students (as well as teachers and other staff) who the program intended to target. For this reason, the evaluation also looks at trends in outcomes for Black boys in grades targeted by AAEEBB programs.⁷ Among the latter group would be included Black boys who were directly affected by their teachers' participation in PLD programs but also Black boys who may not have been affected in any way by AAEEBB program activities. For these reasons, the evaluation team considers outcomes for the mentorship group to be the strongest evidence in favor of program impact. Nevertheless, understanding how their outcomes compare to the broader population of Black boys in AAEEBB schools was also expected to be informative.

With a plan in place to evaluate outcomes for the groups of students expected to be impacted by AAEEBB programming, the evaluation team also sought to identify a comparison group of students. Evidence that outcomes were improved for a particular group, while important, is insufficient for the purposes of program and policy evaluation (Angrist and Lavy, 2008). The most straightforward comparison group of students is non-Black boys (and non-mentorship) in AAEEBB program grades. However, this group of students could still be affected by AAEEBB programming if their teachers were participants, or they could be affected indirectly by their peers' participation in the programs. Therefore, a group of Black boys in program grades in different (but similar) schools is also identified. Outcomes for these four groups are expected to provide the strongest evidence of impact.

Dataset Construction

Student- and school-level data used to compare characteristics of program participants and to estimate the impact of AAEEBB programs are drawn from MSDE's multi-year Education Data Warehouse (EDW). Appendix C describes key measures drawn from the EDW to carry out this evaluation, which are aligned with the evaluation framework. A list of state-assigned student identification numbers for students participating in schools' mentoring and/or Rites of Passage programs was collected as of January 31, 2023. The 13 schools implementing programs in response to

⁷ Mentorship students were identified by schools using a mid-year data collection (see dataset construction section), while program grades were identified using the school survey and the question "in which grades did your AAEEBB programs work to improve outcomes for Black boys?"

recommendations 1.3 and 1.4 submitted 400 student IDs for program participants.⁸ For both mentorship programs and schools' overall AAEEBB programs, we compare student outcomes to a group of Black boys in the same grades as AAEEBB students but who were in a comparison set of schools identified using a propensity score matching procedure. See Appendix D for details on matching. The following table describes the dataset.

Table 2 – Dataset Description for Impact Evaluation

School	Student	Number of Students	Description
AAEEBB	Mentorship	387	Students in AAEEBB mentorship/Rites of Passage programs.
AAEEBB	Black boy	1,626	Black boys in program grades who were not in mentorship programs.
AAEEBB	All other	5,902	All other students in program grades who were not in mentorship programs.
Matched	Black boy	2,086	Black boys in program grades.

Note: Dataset is based on sample of students that from the September 30 enrollment collection.

Analysis

The primary analysis for the impact evaluation relies on a difference-in-differences design. This approach compares the year-to-year trend in an outcome measure for the intervention and comparison groups. If the trend shows an improvement for AAEEBB students relative to the comparison group of students, this is assumed to represent evidence that schools' AAEEBB programs, on average, improved outcomes for students. To increase the plausibility of this assumption (in other words, to rule out additional alternative explanations) and to estimate uncertainty in estimates of impact, the simple trends comparison is complemented with a regression analysis. One important step in the regression analysis is to use available data to find Black boys in similar schools who share characteristics with mentorship students, and these two groups are the only ones who enter this stage of the analysis (in other words, AAEEBB Black boys who were not in mentorship programs and all other students in AAEEBB schools are dropped from the regression analysis). Appendix E describes additional details about the regression analysis. While a regression analysis improves on a simple descriptive trends analysis in important ways, it may not fully account for a key assumption of the difference-in-differences design; namely, that outcome trends would have moved in parallel between mentorship students and similar matched school Black boys. Therefore, an important step in measuring program impact is in thinking through other factors that could have driven outcome differences between these two groups of students.

⁸ Of the 400 IDs submitted, 13 were for students who did not appear in AAEEBB schools as of 9/30.

Process Evaluation Findings

EQ 1: WHAT ARE THE CHARACTERISTICS OF STUDENTS AND SCHOOLS PARTICIPATING IN AAEEBB PROGRAMS?

Schools in year two were the same 14 pilot schools that were selected for year one, chosen based on their local superintendent's stated interest in participation. Table 3 shows characteristics of the 14 schools (column 1) compared to all non-AAEEBB Maryland Public Schools (column 2) for the 2021-2022 school year. The table shows there are some important differences between AAEEBB and other Maryland Public Schools. For example, school climate was significantly lower for both students and educators in AAEEBB schools. Average assessment scores for Black boys were also significantly lower in AAEEBB schools, and Black boys in AAEEBB schools were suspended at higher rates.

Table 3 – School Characteristics by AAEEBB Status, 2021-2022 School Year**Panel A: Students**

Characteristic	AAEEBB	Non-AAEEBB	N
Prop. Student w/ Disabilities	0.13	0.12	1264
Prop. English learner	0.10	0.12	1264
Prop. Economically Disadvantaged	0.32	0.28	1264
Prop. Asian	0.04	0.06	1264
Prop. Black or African American	0.44	0.32	1264
Prop. Hispanic/Latino of Any Race	0.20	0.20	1264
Prop. White	0.25	0.36	1264

Panel B: Teacher

Characteristic	AAEEBB	Non-AAEEBB	N
Prop. teachers Asian	0.03	0.04	1262
Prop. teachers Black or African American	0.24	0.18	1264
Prop. teachers Hispanic/Latino of Any Race	0.04	0.04	1262
Prop. teachers White	0.67	0.72	1262
Prop. teachers with post-grad degree	0.69	0.71	1262
Avg. teachers experience (years)	11.65	12.42	1262

Panel C: School characteristics related to evaluation framework

Characteristic	AAEEBB	Non-AAEEBB	N
Avg. climate score student	4.10	5.39**†	1218
Avg. climate score educator	4.88	6.26***†	1245
Avg. Black boys MCAP 2022 Score Std. (ELA)	-0.56	-0.34*	1264
Avg. Black boys MCAP 2022 Score Std. (Math)	-0.51	-0.28*	1264
Prop. Black boys suspended	0.14	0.06**†	1264
Avg. Black boys' days removed during suspension	4.06	3.45	1240
Avg. Black boys' attendance rate	0.90	0.89	1243
Prop. Black boys' failed ELA course	0.06	0.05	1196
Prop. Black boys' failed Math course	0.05	0.05	1191

Note : * p<.05 ** p<.01 *** p<.001. Two-sided p-values are estimated using t-tests of mean differences. There are 14 AAEEBB schools. Non-AAEEBB schools include all public schools, but do not include vocational-technical, special education centers, alternative educator centers or public charters, to provide an appropriate comparison for the 14 AAEEBB schools.

† P-value remains statistically significant at the .05 level after applying a multiple hypothesis adjustment.

Schools were required to select students and teachers for participation in their AAEEBB programs. Table 4 shows the average characteristics of mentorship students, non-mentorship Black Boys and all other students in AAEEBB schools. The sample of students is restricted to those in grades in which schools' AAEEBB programs were designed to improve outcomes for Black boys (based on their response to this question in the school survey) and does not include students in other grades. Panel B shows that mentorship students, on average, were significantly different from other students in AAEEBB schools on most characteristics examined.

Table 4 - Characteristics of Students in AAEEBB Schools, 2022-2023 Academic Year

Panel A: Demographics

Characteristic	Mentorship students	Non-mentorship Black Boys	All other students
Male	1.00	1.00	0.35
Asian	0.01	0.00	0.07
Black/African American	0.86	1.00	0.30
Hispanic/Latino	0.06	0.00	0.26
White	0.02	0.00	0.29
Other race/ethnicity	0.06	0.00	0.07

Panel B: Service groups and prior year outcomes

Characteristic	Mentorship students	Non-mentorship Black Boys	All other students
Economically Disadvantaged	0.54	0.41	0.37†
Students with Disabilities	0.14	0.20	0.11†
English learner	0.06	0.02	0.11†
MCAP ELA (Std.)	-0.70	-0.54	-0.10†
MCAP Math (Std.)	-0.58	-0.43	-0.16†
Course Failure ELA	0.06	0.04	0.03†
Course Failure Math	0.04	0.04	0.03†
Absence Rate	0.08	0.09	0.09
Suspension	0.16	0.11	0.05†

Note : Sample of students is restricted to those in AAEEBB program grades (for both groups of schools). All characteristics are measured as proportions unless otherwise indicated. † = difference between value and value for mentorship students remains significant at the .05 level after adjusting for multiple hypothesis testing (only characteristics in Panel B are included in adjustment). There were 387 mentorship students, 1,626 non-mentorship Black boys, and 5,902 “all other” students in 2022-2023 (see Appendix D for more information about sample counts).

Summary:

- 14 schools, chosen by LEA superintendents, continued implementing AAEEBB programs in year two. These schools had lower perceptions of school climate and higher rates of suspension for Black boys, compared to other Maryland public schools based on data from 2021-2022. These differences are statistically significant. AAEEBB schools also had higher proportions of Black/African American students and lower proportions of White students, had teachers with fewer years of experience, and Black boys in AAEEBB schools scored lower on state standardized assessments. Although these differences are not statistically significant, they may be considered practically meaningful.
- Students participating in AAEEBB programs were more likely to be economically disadvantaged compared to both Black boys and non-Black boys in program grades. Participants also had lower state standardized assessment scores and equal or higher rates of course failure, as well as higher rates of suspension. These differences were all found to be statistically significant.

EQ 2: WHAT WERE SCHOOLS' GOALS AND OUTCOMES FOR THEIR AAEEBB PROGRAMS IN YEAR TWO?

This section draws from the year two AAEEBB school survey to understand how schools set goals and whether they achieved outcomes for their AAEEBB programs. Table 5 summarizes data on whether schools achieved their goals for the year. Schools were asked to list up to five goals that were part of their school's implementation plan for the year. They were then asked whether their school had achieved the goal and to provide the data points they used to evaluate their goal.

Table 5 – Goals and Outcomes for AAEEBB Programs

Goals	Total across schools	Avg. per school
Number of goals submitted	42	3.0
Number of goals achieved	30	2.1
Number of goals not achieved	4	0.3
Number of goals N/A*	8	0.6

Note: Schools were instructed to report N/A if they had not yet received data to evaluate a goal.

In addition to an overarching look at whether schools achieved their goals as they had proposed them, MSDE worked with schools to improve the clarity and specificity around one to two goals that were submitted as part of their school's implementation plan in the fall of 2022. Schools were encouraged to refine their goals to be specific (related to time periods, academic subjects, data sources), measurable (targets should be quantified where possible), and to measure change over time. Schools were also encouraged to consider factors such as whether their outcome measures were appropriate in relation to the objective of their programs, and the period over which goals would be evaluated. Lastly, schools were encouraged to identify a comparison group for AAEEBB participants to track progress over time. Although most implementation plan goals simply aimed to measure whether an outcome (e.g. test scores) improved for AAEEBB students, MSDE encouraged schools to measure a change in outcomes for participants in relation to a group of non-participants. The objective, in this case, was for schools to be able to better attribute the outcomes they observed to the AAEEBB program, rather than another factor.

In total, 17 out of 18 approved goals for the 14 schools were submitted as part of the year two schools survey. Out of the 17 goals, 11 provided data on AAEEBB program participants, but only three goals also included data for a relevant comparison group. Even among the 11 goals that did provide data, however, there was often a lack of clarity to be able to assess whether there was improvement observed for AAEEBB participants. Findings related to schools' goal setting and use of data are discussed further in the Conclusion and Recommendations section.

Summary: On average, schools reported achieving two out of every three goals they had for their programs in year two. While this finding is positive, schools should continue to strengthen their approach to measuring and using data to be able to more convincingly show it is their AAEEBB programs, and not other programs or initiatives, that are improving student outcomes.

EQ 3: WHAT WERE THE CHARACTERISTICS OF SCHOOLS' AAEEBB PROGRAMS?

In this section we draw on questions from the year two school survey to answer questions about the characteristics of schools' AAEEBB programs. Although no two AAEEBB programs look the same, we grouped programs into mentorship, professional learning/development programs, and curricular/academic supports.

Mentoring/Rites of Passage Programs

In total, 13 schools implemented a mentoring or Rites of Passage program. AAEEBB mentorship programs took many forms but included regularly scheduled meetings and other activities focused on providing academic, social emotional and other non-academic supports to participants. Mentors were selected from among community members and/or partnerships with outside organizations, were teachers and/or non-teaching staff, and may have included older boys, such as high school students. While the content of mentorship meetings may have included academic supports such as help with homework and/or setting academic goals, the mentorship sessions were often described as a space to experience positive affirmations, and a safe space for group and/or self-reflection. Several schools used mentorship sessions to promote leadership skills, financial literacy, and/or etiquette training. Lastly, most schools supplemented regular mentorship sessions with a host of other activities, such as field trips focused on academic and/or non-academic content, community and parent engagement activities, and speaker visits. One school implemented a Rites of Passage program, which included a ceremony focused on teamwork, community building and fostering cooperation among participants.

A few schools stated they faced challenges in maintaining regular participation of mentors and/or other leaders. One school reported plans for adopting a formal application in order to "promote students' motivation in the program." Other schools described positive aspects of their program, such as that they provided an opportunity to engage with adults around an art form or some field of expertise. It was also clear that some schools engaged in experimentation to find out what was perceived to work and what did not. For example, one school had to experiment with different formats for mentorship meetings, including when and where the meetings would take place, before deciding on one that worked for everyone.

One important factor in understanding schools' mentorship programs is how mentors are selected and trained. Schools were asked a series of multiple response questions about how mentors were selected and trained, and Table 6 summarizes this information. Schools were allowed to select more than one option. In total, in eight schools (62% of the 13 schools that implemented mentorship programs) mentors were selected by school staff. In seven schools (54%) mentors were volunteers, and in six schools (46%) mentors were selected through partnership with an outside organization. In only three schools (23%) were mentors selected through a formal application process. Additionally, schools were

asked which factors were considered in selecting mentors. Twelve (92%) of the 13 schools selected mentors based on prior knowledge/familiarity with students or mentees. Additionally, nine (69%) schools selected based on demonstrated interest, and seven (54%) selected based on prior qualifications. Although these categories are not mutually exclusive, the data provide some information to understand how mentors were selected into the program.

Table 6 – Selection and Training for Mentors

How were mentors/role models selected for the program?*	N	%
Through partnership with an outside organization	6	46
Through a formal application process	3	23
They were selected by school staff	8	62
They were volunteers	7	54
Other	0	0

Which factors were considered when selecting mentors/role models?*	N	%
Prior qualifications	7	54
Prior knowledge/familiarity with students/mentees	12	92
Demonstrated interest	9	69
Unsure	0	0
Other	1	8

Did your school experience challenges in recruiting qualified mentors/role models?	N	%
Yes	6	46
No	5	38
Unsure/not applicable	2	15

Was training provided for mentors/mentees?	N	%
Yes	8	62
No	3	23
Unsure	2	15

Note: *Schools were allowed to select multiple response options. There were 13 schools that implemented mentorship and/or Rites of Passage programs.

Schools were also asked about whether they experienced challenges in recruiting qualified mentors/role models. Given the difficulties in finding mentors for programs during year one, it was important to understand to what extent schools were still experiencing these challenges in year two.

The table shows, echoing what was reported in schools' open-ended responses, that about half (six or 46%) of schools reported experiencing challenges, while five (38%) reported that they didn't experience challenges. Lastly, schools were asked whether training was provided for mentors/mentees. Eight of 13 schools (62%) reported there was training provided for mentors.

Another important factor in understanding schools' mentorship programs is features of engagement between mentors and mentees. Schools provided the number of mentors in terms of their professional role. Table 7 summarizes this information. Schools selected mentors with a range of professional roles, with many schools including a mix of teachers, non-teaching staff and non-staff adults. Twelve (92%) schools reported using teachers, 11 (85%) reported using non-staff adults such as volunteers and community mentors, and 10 (77%) reported using non-teaching staff. A smaller number of schools (five or 39%) reported using older peers as mentors. For the entirety of the AAEEBB program the ratio of mentees to mentors was 1.8, which ranged from 0.6 to 7.5 across schools.

Table 7 – Professional Roles of Mentors

Role	Schools reporting any		Total across schools
	N	%	
Teachers	12	92	48
Non-teaching staff	10	77	35
Non-staff adults (volunteers, community mentors, etc.)	11	85	59
Older peers (students)	5	39	71
Other	2	15	7

Note: There were 13 schools that implemented mentorship and/or Rites of Passage programs

Schools were also asked to report on the frequency of engagement in mentorship compared to other activities that were part of their mentorship programs, along with when during the school day mentees engaged with mentors. Table 8 summarizes this information. Most schools reported that mentees engaged with mentors/older peers serving in a mentorship role either once a week (eight of 13, or 62% of schools) or less than once a week but more than once a month (three, or 23% of schools). On average, students engaged in other types of activities such as field trips, speaker series, and cultural enrichment, less often. Four of 13 (31%) schools reported that students engaged in these types of activities once a month, three (23%) reported less than once a month but more than once a week, and three (23%) reported once a week. Lastly, schools were asked to report when during the school day students/mentees engaged with mentors or older peers serving in a mentorship role. Schools were allowed to select multiple response options. Eight (62%) schools reported this happened during lunch or another break, seven (54%) reported this happened before or after school, five (39%) reported during a regular class period, and four (31%) reported on weekends or holidays. Interestingly, six schools (46%) reported this happened during other times.

Professional Learning and Development Programs

In total, 11 schools implemented professional learning and development programs in response to Task Force recommendations 1.1, 2.4, and/or 3.1. It should be noted that these Task Force recommendations have different intents, and given that schools typically implemented multiple recommendations, professional learning and development (PLD) programs could take varying forms even in the same school. Therefore, the survey form combined open ended and multiple-choice response options to best

understand the content and process of implementing PLD programs. Survey responses indicated that schools' PLD sessions ranged from intensive, single day trainings to monthly or bimonthly virtual trainings to speaker series. Most schools partnered with nonprofit or other community organizations that possessed expertise in areas such as social emotional learning, conflict de-escalation, and/or implicit biases in school settings, and had specific knowledge in the education and life experiences of Black boys. Lastly, a small number of schools mentioned scheduling issues that impacted the delivery of PLD programs.

Table 8 – Engagement in Mentorship Programs

How frequently did mentees engage with mentors/older peers serving in a mentorship role?	N	%
More than once a week	1	8
Once a week	8	62
Less than once a week but more than once a month	3	23
Once a month	1	8
Less than once a month	0	0
Never	0	0

How frequently did students engage in other types of activities (field trips, speaker series, cultural enrichment, etc.)?	N	%
More than once a week	2	15
Once a week	3	23
Less than once a week but more than once a month	1	8
Once a month	4	31
Less than once a month	3	23
Never	0	0

When did students/mentees engage with mentors/older peers serving in a mentorship role?*	N	%
During a regular class period	5	39
During lunch or another break	8	62
Before or after school	7	54
On weekends or holidays	4	31
Other	6	46

Note: *Schools were allowed to select multiple response options. There were 13 schools that implemented mentorship and/or Rites of Passage programs.

One specific factor in understanding schools' PLD is how teachers engaged with these programs. The survey form included multiple choice questions asking schools to report on how often they

implemented or facilitated PLD sessions as part of their AAEEBB programs, and what percent of PLD sessions were required to be attended by participants. Table 9 summarizes the results. It shows that most schools (six out of 11 implementing PLD programs, or 55%) implemented PLD sessions once a month. Three schools (27%) implemented sessions less frequently than once a month, and two schools implemented sessions less than once a week but more than once a month. There were also differences in the extent to which schools required staff/student participation in PLD sessions, with four schools (40%) requiring participation 50% of the time or less, and six schools (60%) requiring participation greater than 50% of the time. Four schools (40%) required full participation, and two schools (20%) had full voluntary participation across PLD sessions.

Table 9 – Engagement in Professional Learning and Development Programs

How often did your school implement or facilitate PLD sessions?	N	%
More than once a week	0	0
Once a week	0	0
Less than once a week but more than once a month	2	18
Once a month	6	55
Less than once a month	3	27
Never	0	0

Approximately what percent of PLD sessions were required to be attended by student/staff participants?	N	%
0	2	14
1-25	1	7
26-50	2	14
51-75	2	14
76-100	4	29
Other	0	0

Note: There were 11 schools that implemented PLD sessions in response to Task Force recommendations 1.1, 2.4, and/or 3.1.

Another important factor in understanding schools' PLD sessions is the content of PLD sessions. Table 10 summarizes responses to questions about the content of AAEEBB PLD programs, including virtual or in-person format, when sessions were offered during the school day, and whether sessions provided an opportunity to practice lessons learned. The table shows that schools implemented a combination of virtual and in-person opportunities in their PLD sessions. Two schools (18%) implemented sessions as solely virtual, four schools (36%) implemented sessions as solely in-person, and 5 schools (45%) implemented a combination. The table also shows that most schools (nine of 11, or 82%) offered PLD sessions before or after school; only one school each provided sessions during a regular class period and on weekends/holidays, respectively. Four schools indicated they offered PLD sessions during another time. Lastly, seven schools (64%) indicated PLD sessions included opportunities to apply lessons learned in practice, while two (18%) said they did not and two (18%) said they were unsure.

Table 10 – Structure of Professional Learning and Development Programs

How were PLD sessions offered?	N	%
Solely virtual	2	18
Solely in-person	4	36
Combination virtual and in-person	5	45
Other	0	0

When were PLD sessions offered?*	N	%
During a regular class period	1	9
During lunch or another break	0	0
Before or after school	9	82
On weekends or holidays	1	9
Other	4	36

Did PLD sessions provide participants with time to practice lessons learned with students and/or other adults?	N	%
Yes	7	64
No	2	18
Unsure	2	18
Other	0	0

Note: *Schools were allowed to select multiple response options. There were 11 schools that implemented PLD sessions in response to Task Force recommendations 1.1, 2.4, and/or 3.1.

Curricular/Academic Supports

A total of five schools implemented curricular/academic support programs for students in response to recommendations 3.2, 3.3, and/or 3.6. These included an eight week STEM program, structured literacy supports using Heggerty, Orton-Gillingham and Really Great Reading programs, STEM tutoring, and two schools that offered single-gender courses for selected students. The content of single-gender classes varied, but one of the courses blended state curriculum standards with students' own family trees and historical backgrounds. Schools did not mention any challenges when implementing these programs when provided the opportunity. One school stated AAEEBB funds helped their school invest in training and materials in support of their science of reading initiatives. The survey included two multiple-choice questions for schools implementing "one-on-one or small group academic support (including gender-specific classes)" in response to recommendations 3.2, 3.3, and/or 3.6, but only two of the five schools indicated they were implementing such programs.

Summary:

Thirteen of 14 schools implemented mentorship or Rites of Passage programs in response to Task Force recommendations 1.3 and/or 1.4, 11 schools implemented PLD programs in response to Task Force recommendations 1.1, 2.4, and/or 3.1, and five schools implemented curricular/academic support programs for students in response to recommendations 3.2, 3.3, and/or 3.6.

- With regard to mentorship programs, the ratio of mentees to mentors was 1.8 across the 13 schools. In nine schools, mentorship sessions were held weekly or more frequently, while in seven schools, other types of activities (such as field trips, speaker series, and cultural enrichment) occurred once a month or less frequently. Most schools held mentorship sessions during a scheduled break or before or after school, while five schools held mentorship sessions during a regular class period.
- With regard to mentor recruitment and training, 12 schools selected mentors based on prior knowledge/familiarity with students while three schools required a formal application process for mentors. About half (six) of schools experienced challenges recruiting qualified mentors, and more than half (eight) provided training to their mentors.
- With regard to PLD programs, although the content and purposes of PLD sessions varied based on the recommendation, most (eight) schools implemented PLD sessions once a month or more frequently, and roughly half (six) required participation at more than 50% of sessions. Nine schools offered PLD sessions before or after school, and a majority (7) provided participants with time to apply lessons in practice.
- With regard to academic support programs, the content of these programs varied widely, from structured literacy programs to STEM tutoring to single-gender courses.

EQ 4: WHAT WERE SUCCESSES, CHALLENGES, AND LESSONS LEARNED?

In discussing successes, many schools made a point to state they felt the AAEEBB funding had a positive impact on their students and staff in year two. Even when expressing frustration or disappointment with a challenge, they were quick to state their appreciation and strong feeling that funds were making a difference. Most schools cited program activities such as mentorship and participation in extracurricular activities as successes and felt there were observed differences between how participants behaved by the end of the school year. A few schools also mentioned that their programs were having an impact on “visibility” and helping to change the reputation of their school among the public.

- “Our successes were large this year! We have seen students go from failing to not having a failing grade. One of our students received over 1 million dollars in scholarships alone, several of our students were accepted into college, we started reading a book, we began writing a book!”
- “The AAEEBB program allowed the participants to develop a strong sense of self. Throughout the school year, these students became more confident in their choices, and showed great levels of maturity. The students in the program began to take their education more seriously and started to take strong responsibility of their grades and attendance.”

Schools also took the opportunity to discuss challenges they faced during year two of their programs. The most cited challenges were related to administrative issues, such as delays in receiving funds and experiencing restrictions on how funds could be spent. One school mentioned that it would have been helpful to have a guide with information on key due dates. Several schools also mentioned challenges

specifically relating to their school's mentorship program. As previously discussed, recruiting, identifying, and maintaining consistent participation among a mentor pool was challenging. One school explicitly stated that mentorship suffered because individuals in key positions of leadership within the program left the school. Other schools mentioned that getting buy-in and participation in their mentorship programs from other groups, such as students and community members, was challenging.

- "Our equity extension meetings did not produce a lot of volunteers as the demands of teaching, even with paid compensation, do not always garner a lot of interest."
- "Maintaining a consistent mentoring pool was a huge challenge. Moving forward, we will identify a pool of mentors and assign based on a criteria (sic) that includes access to male role models or a specific match between mentor and student."

In discussing lessons learned and plans for the following year, many schools hinted at additional challenges with their schools' mentoring programs during the school year. Several schools outlined ways in which they would seek to increase motivation and/or accountability for students in their mentorship programs. A couple of schools will institute an application process for interested students rather than having teachers choose participants. A few schools will begin to track attendance in hopes of setting norms and expectations around participation. A few schools felt that it would be best to match students and mentors based on some kind of specific need, interest match, or prior connection. One school will have a family engagement event at the beginning of the year to encourage participation, and another school plans to allow boys greater power and decision-making in mentorship meetings next year. A couple of schools also hope to engage in additional feeder school and community partnerships, as it is felt that a key success of the program is in getting students exposure to the world beyond the school walls. Finally, one school mentioned that the implicit bias training offered to staff was "triggering" for some.

- "Our AAEEBB efforts have been most effective when they have built upon work that we were already doing. Our participation in the pilot has prompted us to go deeper into the work and tailor our work to better meet the needs of our Black boys."
- "The importance of emphasizing culture and historically accurate accounts that students can connect with in class is extremely essential. When students learn that African history did not begin with the hardships felt during the 1600s, it changes their outlook about themselves and their ancestors".

Summary: In open ended questions about successes, challenges, and lessons learned, schools took the opportunity to provide more detail on their AAEEBB programs, and in particular, their mentorship programs. Schools were overwhelmingly positive about their mentorship programs, with many feeling they had observed improvement among participating students ranging from academic to behavioral and socioemotional. Still, schools were up front about the challenges that many had faced. Recruiting and maintaining a consistent mentor pool were common challenges. It was clear that, like year one, year two was an opportunity for many schools to continue experimenting with different ways of structuring their programs, with a wide range of plans to continue modifying and improving the delivery of programs in the next academic year.

Impact Evaluation Findings

This section presents evidence for whether outcomes improved for participating students in year two. Two types of evidence are provided. The first type of evidence is descriptive, in the form of an outcomes analysis, and represents the simplest and most straightforward analysis of to what extent outcomes improved for different groups of students in AAEEBB and matched schools. A second type of evidence, quasi-experimental, is intended to more convincingly answer the question of whether or not schools' AAEEBB programs caused outcome differences for students in year two. Key to the quasi-experimental analysis is the use of regression to estimate program impact. Regression provides a means to do two things; one, it provides a means to calculate confidence intervals which tell whether a difference in the outcomes analysis is statistically meaningful or more likely to be due to some random chance. And two, it estimates the influence of AAEEBB while adjusting for some characteristics of students and schools that could contribute to differences in outcomes. For example, Table 4 shows that mentorship students were more likely to be economically disadvantaged and less likely to be English learners compared to Black boys in matched schools.

State data for Maryland indicates that economically disadvantaged students and English learners both on average have lower assessment growth scores than non-economically disadvantaged and non-English learner students, meaning factors not related to the AAEEBB program likely influence their change in MCAP scores from year to year. Showing outcome trends for each different group of students (Economically disadvantaged vs. non-ED, English learner vs. non-EL, etc.) would increase the number of comparisons considerably and result in a complex analysis. For simplicity, one estimate of impact is provided for each outcome. See Appendix E for more information on regression-adjusted impact estimates.

EQ 5: WHAT WERE THE OUTCOME TRENDS FOR STUDENTS IN AAEEBB AND MATCHED SCHOOLS IN YEAR TWO?

The analyses below examine trends from the pilot year of 2021-2022 to year two in 2022-2023 for four groups of students. Groups of students are identified according to whether they are in "AAEEBB" or "Matched" schools, and their relationship to the AAEEBB program; "mentor" students are those who participated in AAEEBB mentorship and/or Rites of Passage programs implemented in response to Task Force recommendations 1.3 and 1.4 and "Black boy" students are non-mentorship Black boys in the grades in which schools worked to improve outcomes for Black boys as part of their AAEEBB programs. For each outcome measure, raw trends are shown for each group.

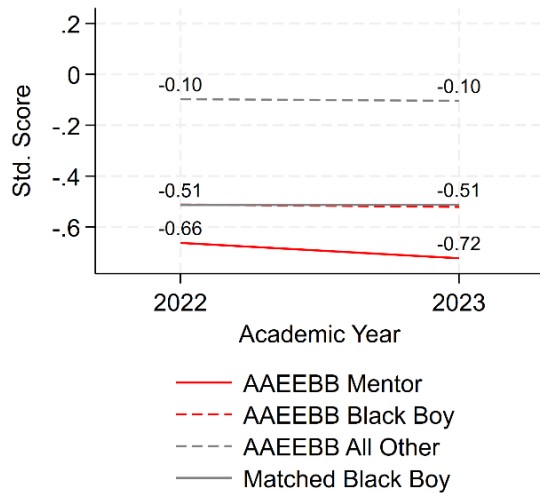
Figure 1 Panel A shows AAEEBB mentor students' ELA scores declined from -0.66 to -0.72, a decrease of 0.06 standard deviations. Both the group of Black boys in AAEEBB schools who were in grades targeted by the programs ("AAEEBB Black Boys") and the group of students in those grades who were not Black boys ("AAEEBB All Other") saw no change in test scores, staying at -0.51 and -0.10 standard deviations between 2021-2022 and 2022-2023 respectively. Lastly, the scores of "Matched Black Boys" also stayed the same between the two school years, at -0.51 standard deviations. Overall, the descriptive evidence does not support that AAEEBB programs improved ELA outcomes. Repeating the same exercise for math (Panel B) suggests that mentor students saw a slight increase in test scores (0.03), though this change was small.

Figure 1 also shows trends in ELA and math course failures for the four groups of students. Panel C shows the rate of ELA course failure for mentorship students declined from 6.2% to 4.0%. This is larger than for other students in AAEEBB schools, and similarly, larger than for Black boys in matched schools.

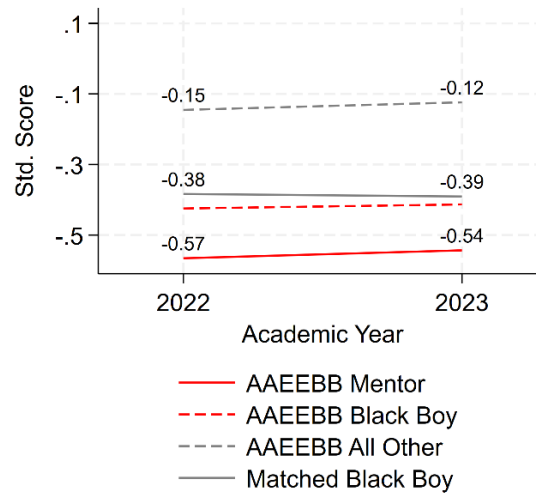
Panel D shows that rates of course failure in math increased for all groups of students in AAEEBB schools, while it declined slightly for Black boys in matched schools.

Figure 1 – Raw Trends in Student Academic Outcomes

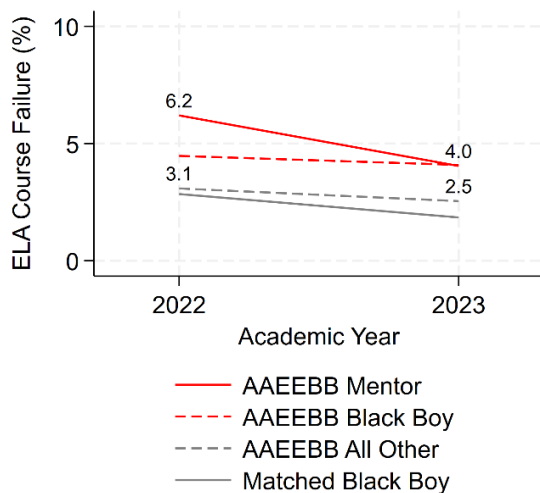
Panel A: MCAP ELA



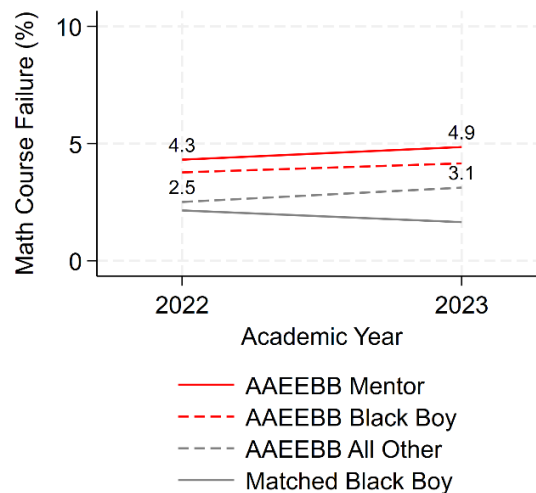
Panel B: MCAP Math



Panel C: Course Failures ELA



Panel D: Course Failures Math

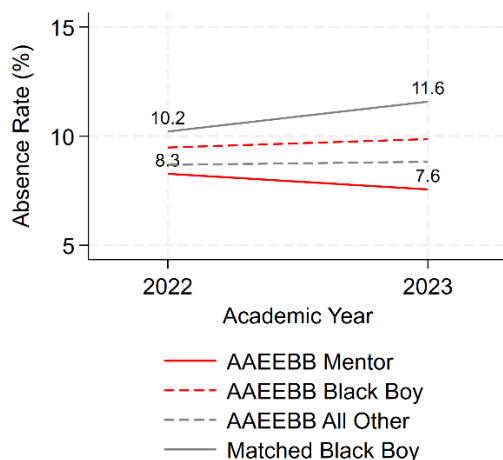


Note: Panels A and B show raw trends in standardized MCAP scores for four groups of students, from 2021-2022 to 2022-2023. The sample of students are those with scores available for both years who were in 4th-8th grade in 2022-2023 (N = 5,516 for ELA and 5,515 for math). Panels C and D show raw trends in course failures for four groups of students, from 2021-2022 to 2022-2023. The sample of students are those with course information available for both years, for each respective outcome (N = 9,483 for both).

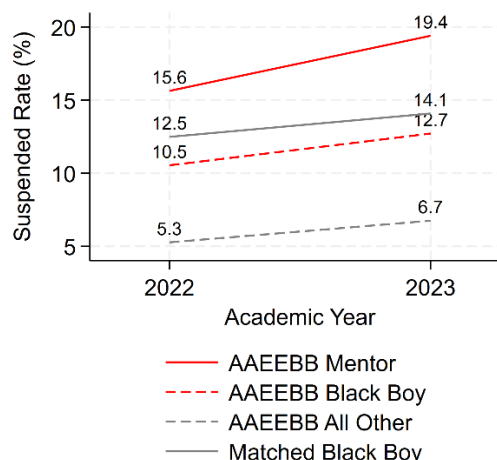
Figure 2 shows trends in absenteeism for the four groups of students, from school year 2021-2022 to school year 2022-2023. Panel A shows that rates of absenteeism declined for mentorship students from 8.3% to 7.6%. The non-mentor “Black Boy” and “All Other” groups saw slight increases in rates of absenteeism, while Black boys in matched schools saw a 1.4 percentage point increase in rates of absenteeism. Panel B shows that rates of suspension increased for all groups on average and the increase was largest (3.8 percentage points) for the AAEEBB “Mentor” group of students.

Figure 2 – Raw Trends in Student Non-Academic Outcomes

Panel A: Absenteeism



Panel B: Suspensions



Note: Panel A shows raw trends in school absence rates for four groups of students, from 2021-2022 to 2022-2023 (N = 9,479). Panel B shows raw trends in suspension rates for four groups of students, from 2021-2022 to 2022-2023 (N = 9,479)

Summary: The outcomes analysis presents mixed evidence of improvement in year two for AAEEBB students. On the one hand, ELA course failures and absenteeism both improved considerably for mentorship students; this is notable, as improvement was either not observed or was less than it was for other groups of students. On the other hand, ELA MCAP scores did not increase for mentorship students by nearly as much as they did for other groups of students, while rates of suspension for mentorship students increased by a wider margin than for other groups of students.

EQ 6: WHAT IS THE IMPACT OF AAEEBB PROGRAMS ON STUDENT OUTCOMES IN YEAR TWO?

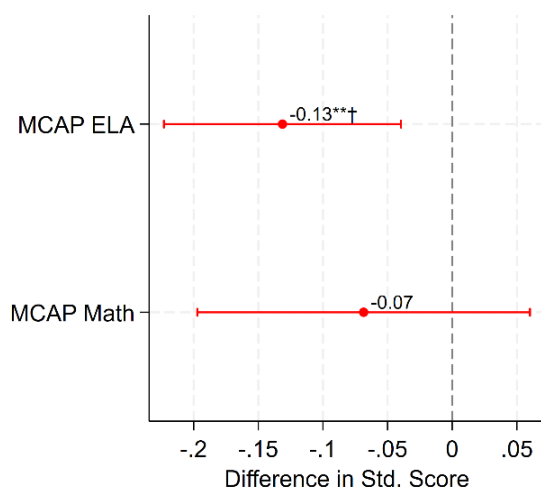
This section presents results of a regression-based impact analysis of AAEEBB program outcomes on mentorship students. To determine impact, a group of Black boys in matched schools who are similar to mentorship students in AAEEBB schools in terms of a number of characteristics, such as their grade level, service group status, and their prior-year outcome measures, are identified. Trends in outcomes from 2022 to 2023 for these two groups of students are then compared, also adjusting for school-level differences. Further explanation of the regression analysis is provided in Appendix E.

Figure 3 presents the results of six regressions, one for each outcome. Panel A shows impact estimates on MCAP scores. There was a statistically significant negative impact of AAEEBB programs on mentorship students' ELA scores, of -0.13 standard deviations. Panel B shows absenteeism was reduced

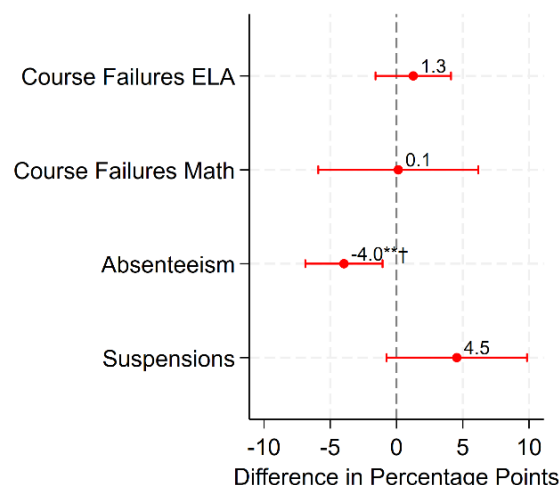
by 4 percentage points for mentorship students compared to similar students. No other program impacts were estimated as statistically significantly different from zero.

Figure 3 – Impact Estimates for Academic and Non-Academic Outcomes

Panel A: MCAP Scores



Panel B: Course Failures, Absenteeism, and Suspensions



Note: * $p < .05$ ** $p < .01$ *** $p < .001$. Figure shows regression-adjusted impact estimates, including 95% confidence intervals, on student-level outcomes, where trends for AAEEBB mentorship students are compared to matched school Black Boys (solid gray lines in Figure 1 and Figure 2). See Appendix E for more information on regression-adjusted estimates. $N = 2,223$ for MCAP ELA and $N = 2,219$ for MCAP Math; $N = 7,878$ for course failures, $N = 7,876$ for absenteeism and suspensions. † P-value remains statistically significant at the .05 level after applying a multiple hypothesis adjustment.

Summary: There is evidence for a statistically significant negative impact (0.13 standard deviations) on mentorship students' ELA assessment scores. There is also evidence that AAEEBB programs reduced the rate of absenteeism for mentorship students by 4.0 percentage points, which is a considerable improvement from the absence rate of 8% in 2021-2022. No other program impacts on student outcomes were observed.

EQ 7: WHAT WERE THE OUTCOME TRENDS FOR AAEEBB SCHOOLS IN YEAR TWO?

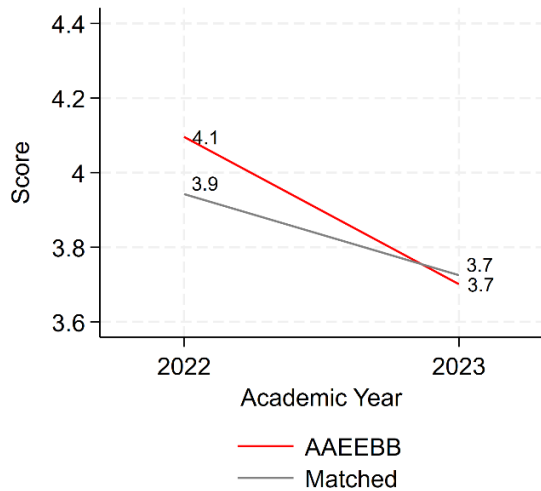
There are only 14 AAEEBB schools, so this section summarizes a simple trend analysis from 2022 to 2023 for AAEEBB and matched schools on the two school climate measures, the proportion of teachers by different race, and race/ethnic disproportionality in suspension rates. The following patterns are observed.

Figure 4 shows trends in school-level measures that are aligned with the evaluation framework. Panel A shows that student climate survey scores declined in both AAEEBB and matched schools but less so in AAEEBB schools from 2021-2022 to 2022-2023. Panel B shows educator climate survey scores in AAEEBB schools stayed at 4.9, while in matched schools declined slightly. Panel C shows that, overall, suspension rates by race followed similar trends for Black boys and Hispanic/Latino boys in AAEEBB and matched schools. Although suspension rates for White boys followed different trends, there's no straightforward reason to believe this had to do with AAEEBB programs. The proportion of Black

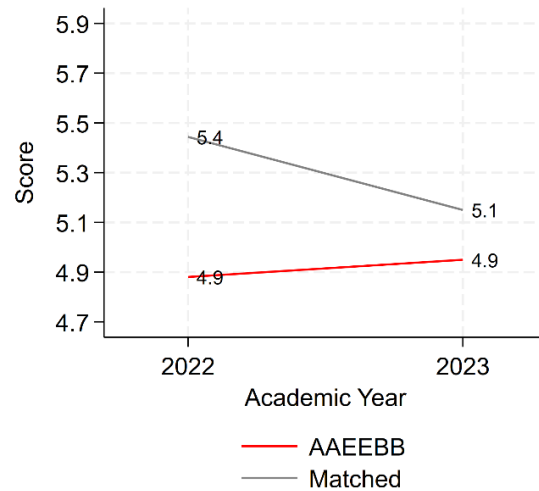
teachers (Panel D) increased slightly in AAEEBB schools but increased more in matched schools. The proportion of White and Hispanic/Latino teachers did not change by much in AAEEBB schools.

Figure 4 – Raw Trends in School Outcomes

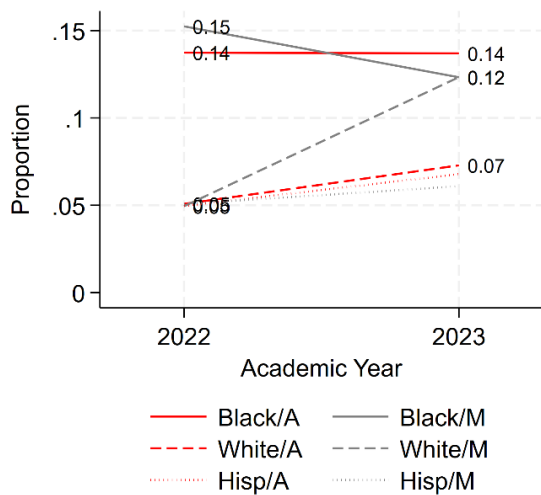
Panel A: Climate (Student)



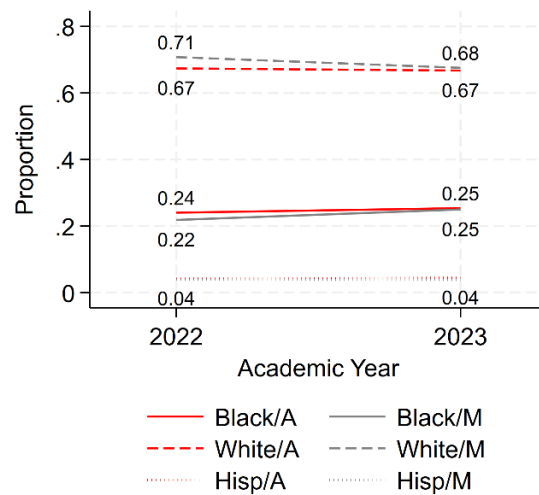
Panel B: Climate (Educator)



Panel C: Suspensions by Race



Panel D: Teacher Race



Note: Panels A through D show trends in school-level outcomes for AAEEBB (A) and matched (M) schools, from 2021-2022 to 2022-2023.

Summary: There is some modest evidence that school climate according to educators improved in AAEEBB schools, but otherwise AAEEBB schools followed matched school trends in most factors related to the evaluation framework.

EQ 8: WHAT WERE THE OUTCOME TRENDS BY SCHOOLS' AAEEBB PROGRAM CHARACTERISTICS?

Data collected in the year two AAEEBB school survey allows for classification of schools' AAEEBB programs along common characteristics. This section presents results of tests of specific hypotheses about the effectiveness of schools' programs. Each hypothesis is tied to specific Task Force recommendations, and for each hypothesis the analysis is for the subgroup of schools that are implementing programs in response to those recommendations. For each hypothesis, schools are grouped according to whether they are implementing programs that have been shown by research to be associated with improved outcomes for students or participants. An unadjusted difference-in-difference estimate comparing trends for mentorship students in the two groups of schools from 2021-2022 to 2022-2023 was calculated. It is important to note that while the hypotheses are based in research, how the hypotheses are operationalized (or measured) in relation to the specific AAEEBB schools is not tied to any specific benchmark, and so each hypothesis includes an explanation of the specific characteristics on which schools were grouped.

Hypothesis 1: Mentorship programs that were of relatively higher intensity improved outcomes for mentorship students to a greater extent than mentorship programs that were of lower intensity.

- There were four schools in which mentors engaged with mentees at least weekly during regular school hours, and where the ratio of mentees to mentors was 2.5 or less.
- Mentorship students in schools implementing high intensity programs had declines in MCAP scores, relative to mentorship students in schools with low intensity programs and had suspension rates that increased. Therefore, there is no evidence that changes in assessment outcomes or course failure rates were better for mentorship students in schools with "high intensity" mentorship programs, compared to mentorship students in schools with "low intensity" programs. There is also no evidence that changes in absenteeism or suspension rates improved for this group of students.

Hypothesis 2: Programs that featured more intensive recruitment and training of mentors resulted in better outcomes for mentorship students, compared to mentorship programs with less intensive recruitment and training.

- There were three schools where mentors were hired through a formal application process and were provided training.
- Mentorship students in schools with more intensive recruitment and training had declines in MCAP scores relative to mentorship students in schools with low intensity programs and had suspension rates that were higher. Similar to mentorship program intensity, there is no evidence for improvement in any of the six outcomes as a result of schools' mentorship recruitment and training practices.

Hypothesis 3: PLD programs that were of relatively higher intensity improved outcomes for mentorship students and/or Black boys in AAEEBB schools to a greater extent than PLD programs that were of lower intensity.

- There were five schools where PLD sessions occurred once a month or more frequently and required attendance for over 50% of sessions.

- There is no evidence that assessment outcomes, course failure rates, absenteeism or suspension rates improved for mentorship students in AAEEBB schools with “high intensity” PLD programs, compared to mentorship students in AAEEBB schools with “low intensity”.

Hypothesis 4: PLD programs where participants had opportunities to practice and apply lessons learned resulted in better outcomes for mentorship students and/or Black boys in AAEEBB schools.

- There were 6 schools where AAEEBB PLD programs featured opportunities to apply lessons learned in practice.
- There is no evidence that outcomes improved for mentorship students on any outcome in schools with these opportunities.

Summary: The collection of data on characteristics of schools’ AAEEBB programs allows for testing of specific hypotheses about whether schools with more intensive and/or arguably more evidence-based AAEEBB programs saw better outcomes for specific groups of students. This analysis does not find consistent evidence in favor of any of the four hypotheses tested.

Analysis of Program Budgets and Expenditures

This section presents an analysis of program budgets and costs provided by schools. Budget information is compiled first by analyzing schools' implementation plans submitted in the fall of 2022. As part of their plans, schools were required to list specific budget items and amounts as part of strategies they outlined to address each Task Force recommendation. As part of the year two school survey administered at the end of the school year, schools were asked to review their implementation plans and note any major changes to the budget.⁹ Similar to the impact analysis, schools are grouped according to whether they implemented mentoring/Rites of Passage programs in response to recommendations 1.3 and/or 1.4; professional learning and development programs implemented in response to recommendations 1.1, 2.4, and/or 3.1; and curricular/instructional programs implemented in response to recommendations 3.2, 3.3 and/or 3.6.

It should be noted there were inconsistencies in how schools reported budget information. For example, some schools grouped transportation and registration/ticket costs associated with cultural events together under a single overarching estimate, while other schools provided this information separately. Additionally, while schools received on average \$52,000 in year two funds, some used year one funds during year two and some listed plans for using year two funds in the 2023-24 school year. Therefore, schools did not spend uniformly the same amounts during the 2022-2023 (year two) school year. Because of inconsistencies in reporting across schools, precise dollar amounts are not reported. Instead, the analysis focuses on high-level patterns observed across schools.

With those caveats noted, the following patterns were observed:

- Mentoring/Rites of Passage programs were the most expensive Task Force recommendations to implement, where they accounted for 55% of total costs reported across schools, followed by professional learning/development programs (27%) and curricular/instructional programs (18%).
- The most substantial costs were related to consultant/partnership (27% of total costs reported) followed by staff compensation (24%), tickets/registration/entrance fees for events (for example, cultural enrichment as part of mentorship programs) (16%), transportation (15%), materials (for example, curricular or technology materials) at 11% and other costs at 6%.
- The proportion of costs devoted to each category varied by type of program.
 - Staff compensation (47%) and materials (25%) made up the largest costs for curricular/instructional programs.
 - Consultant/partnership costs made up 26% of mentoring/Rites of Passage program costs, followed by transportation (24%) and tickets/registration/entrance fees (21%).
 - Consultant/partnership costs made up 42% of professional learning and development costs, while staff compensation made up 27%.

⁹ About half of the schools reported no changes to the budgets presented in their implementation plan. Among schools that did report changes, three reported shifting funds from other items to pay for additional professional development opportunities. Two schools reported challenges in spending their funds due to specific administrative issues, and one school reported shifting funds to pay for transportation and family summit planning.

Summary: The most substantial costs reported by programs were with consultants/organizational partnerships and staff compensation. Fees associated with field trips and transportation were the next largest costs, followed by miscellaneous materials. On average, schools spent the most on their mentorship programs, followed by professional learning and development programs.

Limitations

A key limitation underlying the impact evaluation stems from the challenge in balancing an imperative to measure outcomes for all students and with data availability. For example, standardized test scores are only available for a subset of students, but absenteeism and course data are available for a much larger number of students. The approach in this study is to measure outcomes where they are available, but an alternative approach and one that could be considered for future analyses could be to identify a subgroup of AAEEBB students, or schools, with data available for all outcomes, and to examine impacts for this group. This limitation also has implications for the inability of this study to test key assumptions underlying the approach to estimating the causal impact of the program.

Conclusion

This evaluation report analyzes program implementation and impact for the 14 AAEEBB schools in year two of grant funding. Regarding program implementation, the report examines data on school and student selection for AAEEBB programs, describes program activities and summarizes schools' experiences and lessons learned from the second year. It also provides some limited information on program costs that may be useful to schools planning to implement Task Force recommendations in the future.

With regard to student and school selection for AAEEB programs, state administrative data show that, consistent with the aims of the Task Force, AAEEBB programs are reaching Black boys who face academic and socioemotional challenges in Maryland public schools. In school year 2021-2022 the 14 AAEEBB schools had significantly higher suspension rates for Black boys, and significantly lower perceptions of school climate among both students and educators. Within the 14 schools, students participating in AAEEBB programs were more likely to be economically disadvantaged than were other groups of students, and participating students also had lower state standardized assessment scores, equal or higher rates of course failure, and higher rates of suspension.

AAEEBB schools adopted a variety of Task Force recommendations, but most implemented some combination of mentorship programs along with professional learning and development focused on conflict resolution, anti-bias and/or culturally responsive teaching practices. A smaller number of schools (five of the 13) offered curricular or academic support for literacy, STEM, and/or dedicated single-gender classrooms for specific courses. Data submitted by schools at the end of year two suggest that on the whole, AAEEBB programs were implemented in a way that is consistent with what is known about how to tailor effective programming to support the academic and socioemotional needs of Black boys. For example, mentorship programs had adequate ratios of mentees to mentors and the majority of them met weekly or more frequently. Most schools' PLD sessions were also offered once a month or more frequently, and roughly half required staff participation at the majority of sessions offered. Most schools also appeared to follow evidence-based practices for mentor recruitment, training, and professional learning and development, although it was common for schools to report challenges in recruiting and maintaining a consistent mentorship pool.

To better understand program impact, the evaluation team used state administrative data to examine trends in academic, attendance, and behavioral outcomes for AAEEBB students (focusing on students in mentorship programs) from 2021-2022 to 2022-2023. The impact analysis finds that in year two, AAEEBB students reduced their rate of absenteeism, although their academic outcomes were more mixed. More specifically, the quasi-experimental analysis shows the rate of absenteeism for mentorship students was reduced by four percentage points, which is a considerable improvement from the absence rate of 8% in 2021-2022. At the same time, there was a statistically significant reduction (0.13 standard deviations) in mentorship students' ELA assessment scores. Finally, the analysis does not find evidence that any specific aspects of schools' AAEEBB programs, such as their frequency and/or intensity of participant engagement, was associated with improved outcomes.

Recommendations

The following are recommendations for schools *implementing programs* aligned to Task Force recommendations and meeting the academic and socioemotional needs of Black boys:

- Schools should ensure a continued focus on meeting the academic needs of Black boys in supported programs. This evaluation report suggests in year two schools had mixed success in this regard, with participating Black boys improving on some measures and falling behind in others. At a minimum, schools in the AAEEBB pilot should review their programs to ensure participating students have adequate time both to participate in programs and focus on academics.
- Schools implementing mentoring programs will inevitably require some time to experiment and iterate to build a successful program that meets the needs of Black boys. However, a few lessons can be drawn from the AAEEBB pilot. First, recruiting and sustaining a mentorship pool made up of qualified individuals who also have the knowledge/experience to meet the academic and socioemotional needs of Black boys can be challenging. Relatedly, sustaining the interest and active participation of students can also be challenging. Schools in the AAEEBB pilot experimented with ways to make this process more efficient, including building partnerships with organizations in the community and instituting a more formal application process for both mentors and mentees. Second, many schools in the pilot struggled to settle on a consistent schedule and frequency of mentoring sessions. Thus, schools should give strong consideration at the outset of a program as to how frequently and in what setting (during class time, after school, etc.) mentorship sessions will be offered.
- Schools should combine learning and experimentation with a more thorough understanding of the existing evidence base on what has shown to be effective in terms of mentorship, PLD and academic and socioemotional support for Black boys. Schools should be encouraged to consider “best practices” such as small mentee/mentor ratios and opportunities to apply PLD lessons in practice. Most importantly, schools should understand the success and limitations of similar efforts and how they may be related to their own schools’ choice of programs.

The following are recommendations to aid future *monitoring* efforts in support of schools’ efforts to evaluate program goals and outcomes:

- Improving goals for future efforts should be an important focus. The development of goals and outcomes showed some improvement in year two, but many schools still struggled to write specific, clearly articulated goals. In future evaluations of any program, it should be required to define goals and identify a data source and approach used to evaluate their outcomes. At the end of the school year, rather than an open-ended form for schools to provide information on how they evaluated their goal, schools should be required to enter specific information and data points into a Qualtrics form. This could provide greater assurance that goals both adhere to an appropriate framework for evaluating programs and are submitted with necessary information to objectively evaluate outcomes. While this can improve the transparency and shared understanding of monitoring efforts, it should be remembered that the ultimate objective of this exercise is helping schools to properly evaluate and understand the effects of programmatic efforts.

The following are recommendations for future evaluation work of district- and school-level efforts to implement programs:

- Include a qualitative component in the evaluation to better understand the dimensions of school climate and the daily experiences of students in AAEEBB programs.
- Future analyses should aim to estimate outcomes for a consistent sub-sample of participating students, such as students who have data on all outcomes of interest, including MCAP scores.

REFERENCES

- Bacher-Hicks, A., Billings, S.B., and Deming, D.J. (2019). The school to prison pipeline: Long-run impacts of school suspensions on adult crime. NBER Working Paper 26257.
- Bernstein, L., Dun Rappaport, C., Olsho, L., Hunt, D., and Levin, M. (2009). Impact Evaluation of the U.S. Department of Education's Student Mentoring Program (NCEE 2009-4047). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Darling-Hammond, L., Hyler, M. E., Gardner, M. (2017). Effective Teacher Professional Development. Palo Alto, CA: Learning Policy Institute.
- Dee, T.S. and Penner, E.K. (2021). My Brother's Keeper? The Dropout Effects of Targeted Educational Supports. *Journal of Policy Analysis and Management*, 40(4).
- Fryer, Jr. R. G. 2017. "The Production of Human Capital in Developed Countries: Evidence from 196 Randomized Field Experiments." *Handbook of Economic Field Experiments* 2: 95–322.
- Guryan, J., Ludwig, J., Bhatt, M. P., Cook, P.J., Davis, J.M.V., Dodge, K., Farkas, G., Fryer, R.G., Mayer, S., Pollack, H. Steinberg, L., Stoddard, G. (2023). Not Too Late: Improving Academic Outcomes among Adolescents. *American Economic Review* 113(3).
- Hill, C.J., Scher, L., Haimson, J., & Granito, K. (2023). Conducting implementation research in impact studies of education interventions: A guide for researchers. (NCEE 2023-005). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.
- Institute of Education Sciences. (2013). WWC Review of the Report 'High School Students as Mentors: Findings from the Big Brothers Big Sisters School-Based Mentoring Impact Study'.
- Jackson, C.K. (2018). What do test scores miss? The importance of teacher effects on non-test score outcomes. *Journal of Political Economy*, 126(5).
- Johnson, R.C. and Jackson, C.K. (2019). Reducing inequality through dynamic complementarity: Evidence from Head Start and public school spending. *AEJ: Applied*, 11(4)
- Kautz, T., Tilley, C., Ross, C., and Larkin, N. (2020). Development of a School Survey and Index as a School Performance Measure in Maryland: A REL-MSDE Research Partnership. REL Mid-Atlantic. Available at https://ies.ed.gov/ncee/rel/regions/midatlantic/pdf/REL_MD_School_Survey_2020.pdf.
- Kraft, M.A., and Falken, G.T. (2021). A Blueprint for Scaling Tutoring and Mentoring Across Public Schools. *AERA Open*, 7(1).
- MSDE (n.d.) Task Force on Achieving Academic Equity and Excellence for Black Boys. Transforming the culture of Maryland's schools for Black boys. <https://marylandpublicschools.org/stateboard/Documents/2021/0427/MSDETransformCultureforBlackBoys.pdf>
- Nasir, N.S., and Givens, J.R. (2018). Love Works: Manhood Development Classes and the Pedagogy of Black Male Instructors. In *We Dare Say Love: Supporting Achievement in the Educational Life of Black Boys*. Nasir, N.S., Givens, J.R. and Chatmon, C.P. (Eds.)
- Neitzel, A.J., Lake, C., Pellegrini, M., Slavin, R.E. (2021). A Synthesis of Quantitative Research on Programs for Struggling Readers in Elementary Schools. *Reading Research Quarterly*, 57(1).

Popova, A., Evans, D.K., Breeding, M.E., Arancibia, V. (2022). Teacher Professional Development around the World: The gap between Evidence and Practice. World Bank Researcher Observer, 37(1).

Schochet, P.Z. (2008). Technical Methods Report: Guidelines for Multiple Testing in Impact Evaluations. IES: Washington, DC.

The Abell Foundation. (2021). Literacy Tutoring for Baltimore: What We Know, Where We Are, and How to Move Forward.

TNTP. (2015). The Mirage: Confronting the Hard Truth About our Quest for Teacher Development.

Villavicencio, A., Klevan, S., and Kemple, J. (2018). The Expanded Success Initiative: Challenges and Progress in the Pursuit of College and Career Readiness for Black and Latino Young Men. The Research Alliance for New York City Schools: NYU Steinhardt.

Wei, R. C., Darling-Hammond, L., Andree, A., Richardson, N., Orphanos, S. (2009). Professional learning in the learning profession: A status report on teacher development in the United States and abroad. Dallas, TX. National Staff Development Council.

West, M.R., Kraft, M.A., Finn, A.S., Martin, R.E., Duckworth, A.L., Gabrieli, C.F.O., Gabrieli, J.D.E. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. Educational Evaluation and Policy Analysis, 38(1).

Appendix

APPENDIX A – SUPPLEMENTARY TABLES AND FIGURES

Appendix Table 1 – List of Participating Schools

School Name	LEA
Golden Ring Middle School	Baltimore County
Pikesville Middle School	Baltimore County
Southwest Academy	Baltimore County
Calvert Elementary School	Calvert
JP Ryon Elementary School	Charles
Westlake High School	Charles
Longfellow Elementary School	Howard
Running Brook Elementary School	Howard
Kent County Middle School	Kent
Kent County High School	Kent
Francis Scott Key Middle School	Montgomery
New Hampshire Estates Elementary School	Montgomery
Catherine T. Reed Elementary	Prince George's
Centreville Middle School	Queen Anne's

Appendix Table 2 – AAEEBB Task Force Recommendations

Study Group	Number	Recommendation
Social, Emotional, and Behavioral Supports	1.1	Require de-escalation and other evidence-based intervention and training for all school staff.
	1.2	Include gender equity as part of the Maryland School Report Card, which would give a true measure of equity or lack of equity being achieved for Black boys.
	1.3	Coordinate structured mentoring programs (adult male and/or peer mentoring) tailored to meet the social and emotional learning needs for identified Black boys in grades K-12.
	1.4	Implement a Rites of Passage program for Black boys in Grades 6-12.
Recruiting and Training Skilled, Competent Teachers and Administrators	2.1	Provide financial incentives to recruit and retain racially and ethnically diverse teachers and administrators in the profession.
	2.2	Require school district representatives with teacher and administrator hiring responsibilities (e.g., human resource officers, supervisors, and school principals) to have training on equitable hiring practices of racially and ethnically diverse teacher candidates.
	2.3	Recruit racially and ethnically diverse teacher and administrator candidates into programs to educate and prepare them to enter the profession.
	2.4	Provide all teachers and school-based administrators, along with all other district personnel, continued professional development on culturally responsive teaching practices and methods to support the academic, social-emotional, and developmental needs of Black boys and young men.
	2.5	Use a multi-prong approach to retain Black men and other ethnically diverse teachers and administrators in the profession.

Study Group	Number	Recommendation
Curricula and Instruction	3.1	Provide professional learning to support the implementation of Culturally Relevant Pedagogy/Culturally Responsive Teaching and Anti-bias practices in classrooms and schools.
	3.2	Address ongoing achievement gaps by using the science of reading (systematic phonics instruction, explicit instruction in phonemic awareness, methods to improve fluency, and ways to enhance comprehension) for grades K-3 and beyond for older struggling readers.
	3.3	Address ongoing gaps in math, and continual math decline through advancing grades by using standards-based, real-world math instruction infused with science, technology, engineering, and math (STEM) activities.
	3.4	Use a proven program, such as The Algebra Project, to address persistent gaps in math and the continual decline in math achievement through advancing grades.
	3.5	Expand Equal Opportunity Schools (EOS) and the African American Male Initiative (AAMI) to address reduced enrollment in Advanced Placement courses and gaps in advancement to college and college graduation.
	3.6	Districts encourage individual co-ed schools to create single gender classes across grades or for selected subjects. Districts may also consider approving charter schools with single-gender classes for boys and girls.
	3.7	An annual symposium on teaching and engaging Black boys hosted by the Maryland Department of Education (MSDE).

APPENDIX B – LITERATURE REVIEW

Recent school-based programs targeted toward male students of color have included features such as a focus on culturally relevant instruction, academic tutoring and mentorship, and efforts to improve school climate (Nasir and Givens, 2018; Dee and Penner, 2021; Villavicencio et al., 2018). The African American Male Achievement (AAMA) program in Oakland, CA includes among its core features a focus on culturally relevant instruction by carefully selected adult Black male role models. Students are taught during a regular class period by instructors who undergo an intensive selection, screening, and training process. A quasi-experimental analysis found the program resulted in a statistically significant reduction in the rate of dropout among high school students targeted by the program (Dee and Penner, 2021). A study of the Expanded Success Initiative from NYC found more mixed impacts, with no evidence that it improved academic outcomes (Villavicencio et al., 2018). Other studies of mentorship programs reported difficulties in implementing programs with fidelity (IES, 2013; Bernstein et al., 2009). These programs were found to have mixed effects on student academic and socioemotional outcomes.

While the literature on mentorship programs is rich with descriptive information on the importance of different individual programs and features, convincing research designs that aim to establish causality are less common. For this reason, the evaluation team drew from recent research reviews on personalized tutoring programs. Although mentoring and tutoring have different purposes and often take different forms, there are many similarities. Recent reviews of high-dosage tutoring programs suggest that frequency and duration are among two key aspects of programs that predict their effectiveness (Kraft and Falken, 2021; Abell Foundation, 2021; Neitzel et al., 2021), while there is evidence to suggest that recruitment, selection, and training are particularly important program features (Guryan et al., 2023; Kraft and Falken, 2021).

Reviews of teacher professional development programs suggest there may be an association between length and intensity of professional development activities and student achievement (Chung et al., 2009; Suk-Soon et al., 2007). However, some researchers have noted these reviews rely on a small number of studies with convincing research designs (Popova et al., 2022), and a prominent case study of U.S. school districts suggested the amount of time teachers spent on professional development activities was not associated with improvement in teaching performance (NCTQ, 2015). Darling-Hammond et al. (2017) suggest it is important that professional development programs not be “one-off” events. While there is mixed evidence to suggest that frequency/intensity of programs are important for improving student outcomes, there is stronger evidence to suggest that specific program features are associated with improved outcomes for students. For example, there is evidence to suggest that professional learning and development programs that provide opportunities for participants to apply lessons learned in practice are more effective than programs that place a heavier emphasis on lectures and more passive activities (Fryer, 2017; Chung et al., 2009).

APPENDIX C – ADDITIONAL DATASET DETAILS

Student- and school-level data used to compare characteristics of program participants and to estimate the impact of AAEEBB programs are drawn from MSDE's multi-year Education Data Warehouse (EDW). The following measures are constructed, which align with the evaluation framework:

- MCAP ELA and math scores, measured at the student level, for 2021-2022 and 2022-2023. Scale scores are standardized to have a mean of zero and standard deviation of one within grade, separately for each school year.
- Course failures in ELA and math, also at the school level. Course information is collected at the end of the school year. A student's failure is coded as 0 if they take a full- or school-year term course with the subject "English Language and Literature", and 1 if they take but receive a failing status. A separate variable is constructed for "Mathematics" in the same manner.
- Attendance, measured at the student level, for the 2021-2022 and 2022-2023 school years. Attendance is measured as the number of days a student was absent as a proportion of the total days in membership at a school (days absent plus days attended). No restriction is made based on days in membership.
- Suspensions, measured at the student level, for the 2021-2022 and 2022-2023 school years. Because suspension data is only captured for students with one or more suspensions, a student receives a value of one if they had one or more suspensions in a school year and zero if they had attendance for that school year.
- Demographic and service characteristics of students as of September 30th in the 2022-2023 school year.
- School climate, measured at the school level, for the 2021-2022 and 2022-2023 school years. The Maryland School Survey is administered to all schools annually as part of Maryland's Accountability system and is designed to produce valid and reliable estimates of the extent to which a school's community and physical characteristics promote social-emotional well-being (Kautz et al., 2020). Climate scores, which are produced separately for students and for educators, are expressed on an ordinal scale from 1 to 10.
- The percentage of teachers who are Black, measured at the school level, for the 2021-2022 and 2022-2023 school years.
- Average suspension rate, measured at the school level, for the 2021-2022 and 2022-2023 school years. Technically this measure is a ratio, as it is expressed as the number of suspensions divided by the number of students enrolled as of 9/30.

APPENDIX D – DETAILS ON MATCHING SCHOOLS

The evaluation team used a two-step matching procedure. First, AAEEBB schools should be matched to a set of comparison schools that are similar in important characteristics. Second, within comparison schools, Black boys in the same grades as those in AAEEBB schools should form the sample of students used as the comparison group. Note that for the quasi-experimental impact analysis, this latter group is further restricted using inverse probability weights estimated using observable characteristics of mentorship students (see Impact Evaluation Findings section). Although there are numerous approaches that could satisfy these conditions, the research team opted to use observable characteristics of schools to estimate a propensity score, and then use nearest neighbor matching without replacement.¹⁰ The result is that for each of the 14 AAEEBB schools, a similar comparison school was chosen from among the set of 1,264 public elementary, middle, high, and combined schools that were open in 2021-2022. In addition to requiring that schools match in terms of grade levels offered, the following characteristics were used in the propensity score estimation:

- Overall proportion of Black/African American students (September 30th enrollment)
- Black boys proportion economically disadvantaged (September 30th enrollment)
- Black boys 2022 MCAP scores (English and Math)
- Proportion of Black boys suspended
- Proportion of teachers who are Black
- A full set of missing flags for any schools missing the previous characteristics, with their missing value replaced with the sample mean.

The table below shows characteristics of AAEEBB schools, as well as characteristics of the population of schools that were used in the matching procedure, and the matched schools. Panel A shows the characteristics that were used in the matching algorithm, and Panel B shows a wider number of important school-level characteristics.

¹⁰ This was implemented using the “kmatch” command in Stata.

Appendix Table 3 – Characteristics of AAEEBB and Matched Schools, 2021-2022**Panel A: Characteristics used in propensity score matching**

Characteristic	AAEEBB Schools (1)	Maryland Public Schools (2)	N	Matched Schools (3)	N
Prop. Black or African American	0.44	0.32	1264	0.46	28
Prop. Black boys Direct Cert.	0.49	0.42	1264	0.44	28
Avg. Black boys MCAP 2022 Score Std. (ELA)	-0.56	-0.34*	1264	-0.53	28
Avg. Black boys MCAP 2022 Score Std. (Math)	-0.51	-0.28*	1264	-0.40	28
Prop. Black boys suspended	0.14	0.06**†	1264	0.15	28
Prop. teachers Black or African American	0.24	0.18	1264	0.22	28

Panel B: Other school characteristics not used in matching

Characteristic	AAEEBB Schools (1)	Maryland Public Schools (2)	N	Matched Schools (3)	N
Prop. Asian	0.04	0.06	1264	0.05	28
Prop. Hispanic/Latino of Any Race	0.20	0.20	1264	0.16	28
Prop. White	0.25	0.36	1264	0.27	28
Prop. Special Ed.	0.13	0.12	1264	0.14	28
Prop. ESL	0.10	0.12	1264	0.07	28
Prop. FARMS	0.51	0.43	1264	0.48	28
Prop. Direct Cert.	0.32	0.28	1264	0.32	28
Prop. Black boys Special Ed.	0.21	0.19	1240	0.21	28
Prop. Black boys ESL	0.04	0.03	1240	0.01	28
Prop. Black boys FARMS	0.64	0.57	1240	0.58	28
Prop. Black boys failed ELA course	0.06	0.05	1196	0.05	27
Prop. Black boys failed Math course	0.05	0.05	1191	0.08	27

Characteristic	AAEEBB Schools (1)	Maryland Public Schools (2)	N	Matched Schools (3)	N
Avg. Black boys days removed during suspension	4.06	3.45	1240	3.82	28
Avg. Black boys attendance rate	0.90	0.89	1243	0.88	28
Avg. climate score student	4.10	5.39**†	1218	3.94	26
Avg. climate score educator	4.88	6.26***†	1245	5.44	27
Prop. teachers Asian	0.03	0.04	1262	0.02	28
Prop. teachers Hispanic/Latino of Any Race	0.04	0.04	1262	0.04	28
Prop. teachers White	0.67	0.72	1262	0.71	28
Prop. teachers with post-grad degree	0.69	0.71	1262	0.66	28
Avg. teachers' years of experience	11.65	12.42	1262	10.94	28

Note: * $p < .05$ ** $p < .01$ *** $p < .001$. Two-sided p-values are estimated using t-tests of mean differences. Maryland Public Schools include all public elementary, middle, high, and combined schools (does not include AAEEBB schools). The 14 AAEEBB schools are located in Baltimore County (3), Charles, Howard, Kent, and Montgomery counties (2 each), and Calvert, Prince George's, and Queen Anne's (1 each). The 14 comparison schools are located in Baltimore City and Charles (3 each), Anne Arundel, Montgomery and Queen Anne's (2 each), and Baltimore City and Carroll (1 each). † P-value remains statistically significant at the .05 level after applying a multiple hypothesis adjustment.

Appendix Table 4 – Characteristics of Students in AAEEBB and Matched Schools, 2022-2023**Panel A: Demographics**

Characteristic	AAEEBB Schools			Matched Schools
	Mentorship Students	Non-mentorship Black Boys	All Other	Black Boys
Male	1.00	1.00	0.35	1.00
Asian	0.01	0.00	0.07	0.00
Black/African American	0.86	1.00	0.30	1.00
Hispanic/Latino	0.06	0.00	0.26	0.00
White	0.02	0.00	0.29	0.00
Other race/eth.	0.06	0.00	0.07	0.00

Panel B: Service groups and prior year outcomes

Characteristic	AAEEBB Schools			Matched Schools
	Mentorship Students	Non-mentorship Black Boys	All Other	Black Boys
Economically Disadvantaged	0.54	0.41	0.37†	0.42†
Students with Disabilities	0.14	0.20	0.11†	0.20†
English learner	0.06	0.02	0.11†	0.02†
MCAP ELA (Std.)	-0.75	-0.59	-0.15†	-0.60†
MCAP Math (Std.)	-0.63	-0.48	-0.21†	-0.46†
Course Failure ELA	0.06	0.04	0.03†	0.03†
Course Failure Math	0.04	0.04	0.03†	0.02†
Absence Rate	0.08	0.09	0.09	0.10†
Suspension	0.16	0.11	0.05†	0.12

Note: Sample of students is restricted to those in AAEEBB program grades (for both groups of schools). All characteristics are measured as proportions unless otherwise indicated. † = difference between value and value for mentorship students remains significant at the .05 level after adjusting for multiple hypothesis testing (only characteristics in Panel B are included in adjustment). There were 387 mentorship students, 1,626 non-mentorship Black boys, and 5,902 “all other” students in 2022-2023; there were 2,086 Black boys in matched schools.

Appendix Table 5 – Sample Counts of Students in AAEEBB and Matched Schools, 2022-2023**Panel A: Demographics**

Characteristic	AAEEBB Schools			Matched Schools
	Mentorship Students	Non-mentorship Black Boys	All Other	Black Boys
Male	387	1626	5902	2086
Asian	387	1626	5902	2086
Black/African American	387	1626	5902	2086
Hispanic/Latino	387	1626	5902	2086
White	387	1626	5902	2086
Other race/eth.	387	1626	5902	2086

Panel B: Service groups and prior year outcomes

Characteristic	AAEEBB Schools			Matched Schools
	Mentorship Students	Non-mentorship Black Boys	All Other	Black Boys
Economically Disadvantaged	387	1626	5902	2086
Students with Disabilities	387	1626	5902	2086
English learner	387	1626	5902	2086
MCAP ELA (Std.)	275	1129	3902	1375
MCAP Math (Std.)	266	1155	3917	1411
Course Failure ELA	371	1565	5544	2003
Course Failure Math	371	1565	5544	2003
Absence Rate	371	1565	5541	2003
Suspension	371	1565	5541	2003

Note: Sample of students is restricted to those in AAEEBB program grades (for both groups of schools).

APPENDIX E – ADDITIONAL DETAILS ON ANALYSIS

The impact analysis uses the following ordinary least squares (OLS) regression equation to estimate the impact of AAEEBB for mentorship students on each outcome from 2021-2022 to 2022-2023:

$$y_{ist} = \alpha + M_{is} + \beta 1([1 = YEAR_{2023}] * M_{is}) + \gamma_p + \epsilon_{ist}$$

In this equation, the outcome measure for student *i* in school *s* at time *t* is regressed on a dummy variable for whether the student is in a mentorship program (*M*) interacted with a dummy variable for the 2023 school year. The model includes fixed effects γ_p for the matched-pair group of schools. Each regression is estimated with student-level inverse probability weights, which are calculated using students' 2023 grade level, direct certification status, English learner status and special education status, as well as their 2022 MCAP scores, course failure rates, absence and suspension rates. All models are estimated with missing data flags. Inverse probability weights assign a weight to each student based on their probability of being a mentorship student, which effectively changes the comparison group of “Black boys in matched schools” to be observationally similar to mentorship students. Standard errors are clustered at the student level.

The identifying assumption for estimating a causal impact of AAEEBB is that conditional on individual and school-level characteristics accounted for in the model, no other within-cohort factors are assumed to have driven differences in MCAP scores for students. The inability to draw on comprehensive pre- and post-period data for all students precludes the ability to thoroughly test this assumption, but future analyses of the AAEEBB program may consider ways to identify and more thoroughly defend assumptions behind causal estimates.

When carrying out statistical inference, it is well known that as the number of hypothesis tests for the same or similar populations increases so does the chance of Type I error, or the chance that a finding will be estimated as statistically significant when there is no true effect (Schochet, 2008). Researchers have a variety of multiple hypothesis adjustments they can choose from in order to reduce this risk. To control for the false discovery rate, or the expected proportion of null hypotheses that will be incorrectly rejected, researchers can apply a procedure described in Benjamini, Krieger, and Yekutieli (2001).¹¹ In doing so, it is assumed that the six student-level outcomes examined in this evaluation (assessment in ELA, assessment in math, course failure in ELA, course failure in math, absenteeism, and suspension rate) fall under six outcome domains and within each domain it is necessary to account for one significance test. The results of this analysis are displayed in each of the relevant figures of the body.

¹¹ This multiple hypothesis adjustment is carried out using the “multproc” command in Stata.