Final Impact Evaluation Report

Evaluation of Achievement Mentoring

U.S. Department of Education

Education Innovation and Research Program

June 2025

Submitted by:
The Policy & Research Group
www.policyandresearch.com

8434 Oak St.
New Orleans, LA 70118

107 Spring St.
Seattle, WA 98104

Katherine Lass, MPH, MSW; Sarah Walsh, PhD; Kelly Burgess, MPH; Jeremy Kuperberg Levin, PhD; Eric Jenner, PhD. 2025. Evaluation of Achievement Mentoring (AM). New Orleans: The Policy & Research Group.

ACKNOWLEDGMENTS

This study would not have been possible without the following current and former team members at PRG: Carolyn Kelly, Rebekah Snider, Alexis Mayfield, Sarah Taft, and Alice Juillard. In addition, we are thankful for the Center for Supportive Schools and Dr. Brenna Bry who developed the intervention and worked diligently throughout the study to provide sites and staff with training and technical assistance in program implementation. We would also like to thank Bruce Randel, PhD, from Century Analytics for his sage guidance and interminable encouragement throughout the course of this study. Finally, we express deep appreciation to the study site partners, staff, and youth who participated in this study.

DISCLOSURE

No potential competing interest to report.

This publication was prepared under Grant Number U411C190018, awarded to the Center for Supportive Schools, from the *Office of Elementary and Secondary Education* within the *U.S. Department of Education*. The views expressed in this report are those of the authors and do not necessarily represent the policies of the funders.

EXECUTIVE SUMMARY

The Center for Supportive Schools (CSS) received a *U.S. Department of Education*, Education Innovation and Research (EIR) Program early-phase grant in 2019 to implement and evaluate Achievement Mentoring (AM), an innovative adult mentoring program, which aims to improve student retention and achievement in school. As part of the five-year grant, CSS contracted with The Policy & Research Group (PRG) to independently evaluate the implementation and impact of the AM program on 10th- and 11th-grade students' attendance, achievement, disciplinary, and social and emotional outcomes. The purpose of this report is to present summative implementation and impact evaluation findings from the project.

AM is a school-based mentoring program for students in Grades 4 through 11 that is implemented over two consecutive school years. The present study focused on the implementation of the program with 10th- and 11th-grade students identified as being at risk for dropping out of school before graduation. AM is grounded in Social-Emotional Learning (SEL) and Social Learning Theory and leverages existing resources, such as school staff, to create a supportive environment for the target population that encourages them to set goals, make more positive decisions, and seek help from their mentors. By offering additional support to these students and opportunities to interact with adults in one-on-one settings, the program seeks to mitigate issues, such as disciplinary referrals, poor academic performance, and school disengagement, that are commonly associated with dropping out. CSS hypothesizes that students who are offered the AM program will build meaningful connections with their mentors, develop SEL skills, become engaged in their schools, and have improved school-related outcomes.

PRG conducted a rigorous impact and implementation evaluation of the AM program's effect on 10th- and 11th-grade student outcomes. The impact study utilized an individual-level randomized controlled trial (RCT) where half of eligible 10th-grade students were randomly assigned to be offered the AM program (treatment) or a class-as-usual control condition. Students were eligible for the program if they met one of three risk indicators in the prior year (ninth grade): missing 20 or more days of school, failing a core course, and/or incurring three or more disciplinary infractions. Outcomes used to assess impact included school attendance, credits earned, suspensions, and social and emotional outcomes (e.g., perceived support, self-efficacy, decision-making skills). We estimate program impact within an intent-to-treat (ITT) framework, regardless of their actual exposure to the AM intervention. Impact estimates were calculated using a regression equation that models the outcome of interest as a function of treatment status, and a series of covariates, including the baseline measure of the outcome. The implementation study aimed to explore the extent to which the AM model was implemented as intended at each study site during the three implementation school years (2021–22, 2022–23, and 2023–24) and the amount of programming students assigned to the AM condition received.

KEY FINDINGS

IMPLEMENTATION OF AM

AM is designed to be implemented across two school years, while students are enrolled in 10th and 11th grades. Mentors are instructed by the program developers and their CSS program managers to meet with their mentee at least 20 times, for 20 minutes, each school year to achieve fidelity of program implementation. Schools varied in their overall length of programming and capacity to reach mentees. In all, 13 schools agreed to participate in the study, 12 of which implemented at least some programming. In addition, several schools did not maintain participation in the project for both years of implementation as intended.

Across cohorts and implementation years, schools had an average of 6 mentors assigned to 10 mentees. This ranged widely based on school capacity and the number of eligible students. Mentors are encouraged by program developers to meet with their mentees at least 20 times per school year to achieve fidelity of implementation. Mentors held, on average, between 14 and 15 mentoring sessions with their mentees over the course of a school year. The average number of mentoring sessions held increased, on average, between the first year of implementation (12 sessions) and the second year (18 sessions).

Mentors documented varying degrees of adherence to the AM model during each session held with their mentee. During each session, mentors were most likely to give their mentee praise for instances of engagement with school followed by helping their mentee set a goal for the coming week and establishing a plan to achieve that goal. Mentors were slightly less consistent with interviewing teachers prior to the session, sharing teacher feedback from interviews, sharing the *Weekly Report Form* (WRF) and eliciting student feedback from the WRF.

Individual-level dosage data suggest that students assigned to the treatment condition did not receive the recommended minimum threshold dosage of mentoring (20 sessions per year). A total of 197 students were randomly assigned to be offered AM at the beginning of their 10th-grade year across all three cohorts. Of these, 161 (81.7%) received some level of mentoring after random assignment, whereas 36 (18.3%) did not receive any mentoring. On average, students attended between 10 and 11 mentoring sessions during their first year of participation and between 9 and 10 sessions in their second year, or roughly half the minimum number of recommended mentoring sessions (20) advised by the program developers.

IMPACT FINDINGS

TENTH GRADE

Benchmark findings at the end of the first time point (10th grade) provide some promising support for the hypothesis that offering regular, individualized mentoring to students who meet one or more risk indicators at the start of 10th grade can improve their academic, behavioral, and social and emotional outcomes. Model estimates indicate that students who were offered AM at the start of 10th grade self-reported feeling more supported by adults at school and more confident in their abilities to practice goal-setting, help-seeking, and academic skills. They also self-reported using critical decision-making skills more frequently, and administrative data indicate they were less likely to be suspended during their 10th-grade year. Although findings related to attendance, credit accrual, and academic self-image were not statistically significant, mean outcomes for AM students were consistently greater than their control counterparts.

Subgroup analyses provide some evidence that the program may be more effective at improving administrative outcomes (discipline, attendance, credits) among students in non-rural schools. Specifically, students assigned to AM at urban or suburban schools earned more credits, attended more days of school, and were less likely to get suspended during 10th grade.

ELEVENTH GRADE

Benchmark findings at the end of the second time point (11th grade) provide modest evidence that offering two years of individual mentoring to students identified as at risk for dropping out of school improves academic, behavioral, and social and emotional outcomes. In the full ITT sample for whom we have data, we do not observe significant impacts on administrative outcomes (attendance, credit

accrual, discipline); however, we do see impacts on select social and emotional outcomes at the end of 11th grade. Model estimates indicate that students assigned to AM self-reported feeling more supported by adults at school, identifying as someone who could succeed academically, and practicing critical decision-making skills more frequently at the end of 11th grade, but did not report greater self-efficacy or peer acceptance.

Subgroup analyses provide some evidence that the program improved students' academic and behavioral outcomes in non-rural schools. Specifically, students assigned to AM at urban or suburban schools earned more credits, attended more days of school, and were less likely to get suspended during 10th and 11th grades. This again suggests that the program may be more effective at improving academic and behavioral outcomes in non-rural settings, though given the very small analytic sample size (n = 44) at this time point, subsequent research is necessary to confirm this trend.

CONCLUSION

Results from our evaluation of AM provide promising evidence of the program's potential to improve outcomes for students identified as being at risk of dropping out of high school. Empirical results at the end of 10th grade suggest that offering one year of AM to treatment students led to comparatively higher scores on measures of perceived support, self-efficacy, and decision-making skills, as well as a lower proportion of students getting suspended compared with control students who were not offered mentoring. Although we expected to observe greater program effects after two years, findings are more modest when we examine the effect of AM on 11th-grade outcomes for the subsample of our ITT population.

The program's goal is to help students who are identified as being at risk for dropping out of high school ultimately make it to graduation. We would argue that the results of this evaluation provide promising evidence that offering one-on-one adult mentoring through the AM model could improve the likelihood that students who are identified as being at risk of dropping out before graduation would graduate from high school. Although the findings from our multiyear, cross-state evaluation do not provide strong evidence that a second year of mentoring continues to improve outcomes, taken together, these trends provide early support for the continued investigation of how AM could improve graduation rates for students who meet one or more risk indicators during ninth grade. It is clear that the AM program would benefit from subsequent, larger scale replication studies to continue to refine its logic model and investigate its effect on student outcomes more broadly and through the high school pathway to graduation.

TABLE OF CONTENTS

INTRODUCTION

Through its Education Innovation and Research (EIR) Program, the *U.S. Department of Education* provides competitive grants to applicants with a record of improving student achievement and attainment in order to expand the implementation of, and investment in, innovative practices that are demonstrated to have an impact on improving student achievement or student growth, closing achievement gaps, decreasing dropout rates, increasing high school graduation rates, or increasing college enrollment and completion rates.[1]

The purpose of this report is to present summative findings from a five-year project that implemented and evaluated an innovative high school adult mentoring program, Achievement Mentoring (AM), which aims to improve student retention and achievement in high school. AM is a school-based high school adult mentoring program for 10th-grade students that is designed to improve social-emotional learning (SEL) skills and improve academic and behavioral outcomes. The program pairs 10th-grade students (mentees) with an adult mentor at their school who is responsible for meeting with each of their mentees weekly to support their academic engagement and increase their likelihood of high school graduation.

Funded through a 2019 EIR early-phase grant, the project was a collaborative effort between the Center for Supportive Schools (CSS), the grantee and program developer, and The Policy & Research Group (PRG), the independent evaluator. PRG conducted a multisite, multiyear individual-level randomized controlled trial (RCT) designed to assess AM's impact on student attendance, achievement, disciplinary, and social and emotional outcomes. PRG also conducted a concurrent implementation study aimed at understanding the extent to which AM was conducted with fidelity at each study site. As originally designed, the study team planned to enroll two cohorts of students during the 2021–22 and 2022–23 school years, with a target of 800 students to be enrolled. Recruitment challenges related to the aftermath of the COVID-19 pandemic resulted in a smaller sample size than initially planned. In all, PRG enrolled 393 tenth-grade students from 13 high schools in 5 East Coast states across 3 cohorts during the 2021–22, 2022–23, and 2023–24 school years.

This report provides summative findings from the five-year evaluation of the AM program. We first present an overview of the program, including the theory of change and program model. We then provide an overview of the impact study, including research questions, control group experience, eligibility criteria, random assignment procedures, outcome measures, and data collection and analytic methods. We then describe the final study and analytic samples and present findings and discussion from the fidelity study and impact analyses. Supplemental details are provided in a series of appendices that present a graphical representation of the AM logic model (Appendix A), implementation fidelity study overview and findings (Appendix B), detailed variable operationalization and analytic methods (Appendix C), detailed impact findings (Appendix D), and the results of several sensitivity analyses (Appendix E).

ACHIEVEMENT MENTORING

THEORY OF CHANGE

Despite rising graduation rates for public high school students in the decade prior to the COVID-19 pandemic, disparities in educational attainment have persisted for students of color, low-income, and

[1] For more information on the EIR program, see https://www.ed.gov/grants-and-programs/grants-special-populations/economically-disadvantaged-students/education-innovation-and-research#Home

high-need students (National Center for Education Statistics [NCES], 2024). Similarly persistent across these groups are disparities in school disciplinary infractions, a significant factor in high school dropout rates (NCES, 2022). Research suggests that better engaging students in school and connecting them to supportive adult relationships are avenues to reduce dropout risk and reduce disparities in educational attainment across social groups (Fall & Roberts, 2012; Johnson et al., 2006; Pianta et al., 2012). High school-aged students have starkly unequal access to mentoring, however, widening opportunity gaps (Gordon, 2016). Youth who are at risk of dropping out are less likely to complete high school, attend college, or achieve positive adult outcomes if they never have a mentor in their adolescent years (Bruce & Bridgeland, 2014). Schools can therefore prevent punitive disciplinary outcomes and improve learning conditions by increasing access to positive relationships with school-based adults for students at risk of dropping out.

Dr. Brenna Bry developed AM based on research regarding disparities in high school completion, the importance of engagement during high school, and the importance of mentoring relationships with supportive adults. Through a partnership between Dr. Bry and CSS, CSS provides the professional development associated with implementing AM. AM is a two-year, school-based program designed for 4th- through 11th-grade students that is designed to improve SEL skills (e.g., goal-setting, help-seeking), improve educational mindsets, enhance student engagement, and improve school-related outcomes (e.g., discipline, progressing in school, attendance). The present study focused on the implementation of the program during the 10th- and 11th-grade school years.

AM is a cognitive-behavioral intervention grounded in SEL and Social Learning Theory. Research indicates that, compared to students who do not participate in such programs, students who receive adult mentoring and regular praise have improved SEL skills such as academic self-efficacy, self-concept, and peer acceptance (Dubois et al., 2011). These skills and supports have been shown to foster students' academic achievement and engagement and reduce problem behaviors. In line with the principles of Social Learning Theory, CSS posits that mentors can help students increase their engagement in school through praise of high-engagement activities they observe and showing students that their positive behavior has immediate, positive consequences.[2] The program's hypothesis is that this positive feedback loop will increase student engagement and ultimately lead to improved academic achievement.

The high school model of AM begins for students during their 10th-grade year and continues through the end of 11th grade. This model is designed to intervene with students who meet one or more common risk indicators for dropping out of high school – chronic absenteeism, multiple behavioral infractions, failure of a core class – often referred to as the "ABCs" of early warning systems (Bruce et al., 2011). CSS hypothesizes that through intervention relatively early in the high school trajectory, students who were previously at risk of dropping out can improve their school engagement and reorient themselves toward high school completion.

The AM program leverages existing resources, such as school staff, to create a supportive environment for high school students at risk of dropping out that encourages them to set goals, make more positive decisions, and seek help from their mentors. By offering additional support to these students and opportunities to interact with adults in one-on-one settings, the program seeks to mitigate issues, such as disciplinary referrals, poor academic performance, and school disengagement that are commonly

[2] Social Learning Theory suggests that all actions, thoughts, and feelings are learned, and that new learning experiences can change ingrained behaviors (Bandura, 1977).

associated with dropping out. CSS hypothesizes that students who are offered the AM program will build meaningful connections with their mentors, develop SEL skills, become engaged in their schools, and have improved school-related outcomes.

THE ACHIEVEMENT MENTORING MODEL

The program's logic model, including its key components, is presented in Appendix A. AM requires the integration of three key components that work together to ensure the mentoring program is implemented as intended: (1) a stakeholder team consisting of administrators, faculty, counselors, and other school staff who work together to make programmatic decisions and incorporate the program into the school; (2) professional development through initial training and monthly coaching sessions for mentors; and (3) mentoring activities, including weekly interviews with mentees' teachers, monthly outreach to parents/caregivers, and weekly mentoring sessions during which mentor-mentee pairs review weekly school reports and set personal goals.

The AM model outlines clear expectations for training, program support, and student-level data collection to support program setup and implementation. AM stakeholder teams at each school select mentors, hold planning meetings, and implement the program with support from CSS. These teams are led by a stakeholder team coordinator (STC) who assists with data collection and acts as a liaison between the school and CSS. CSS provides initial and ongoing professional development to mentors throughout the two-year intervention, including an orientation session, three days of structured training, and ongoing coaching meetings and technical assistance. Each of these professional development activities can be provided virtually or in person as needed.

The AM curriculum consists of weekly, structured mentoring sessions. Students assigned to receive the intervention attend 20-minute pull-out mentoring sessions, once per week and are typically pulled from lunch, physical education, advisory, or another noncore academic block. CSS program managers, who act as technical supports to the school, emphasize that a minimum of 20 mentoring sessions are expected to be offered over the school year for a mentor to meet the minimum fidelity requirement. Before each weekly meeting, mentors complete a *Weekly Report Form* (WRF) based on consultation with one of the mentee's teachers about the student's academic performance and behavior during the week prior. Comments may include references to on-time arrival to class, having materials for class, completing classwork, completing homework, etc. Mentors also solicit a positive or encouraging comment from the teacher that can be shared with the mentee. The teachers with whom mentors hold these consultations are up to their discretion; however, mentors are encouraged to reach out to different teachers throughout the year. During weekly sessions, mentors follow a multistep sequence involving a review of the WRF, student praise, a review of teacher observations, motivational interviewing, and weekly goal-setting.

As part of the program, mentors are supposed to contact each mentee's parent or guardian once per month to share something positive about the mentee's behavior or interactions at school. Through this practice, AM provides a structure in which adolescents are noticed and praised directly and consistently for any actions they have taken to increase their school engagement. The program logic model suggests that praising mentees' positive behavior should influence their concept of self-efficacy and positive feelings about school, creating a positive feedback loop that nurtures longer-term academic success.

Other program activities include mentor-mentee events held at the beginning and end of the school year. Though the mentoring component of AM can be tailored to meet the needs of a particular school,

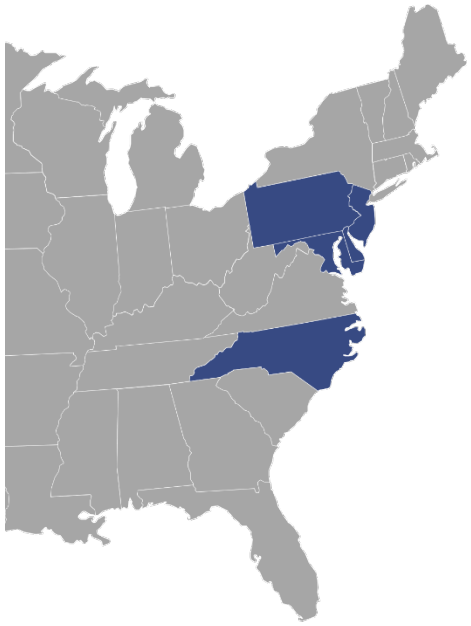
typically, the program begins in the second or third month of the mentees' 10th-grade year with the beginning-of-year kickoff luncheon where mentees are officially introduced to their mentors and provided an overview of the program. This luncheon is organized by the school's STC. In addition, each year of programming concludes with a second mentor-mentee luncheon, also organized by the STC. During these end-of-year luncheons, mentees celebrate their work in AM, identify future goals, and learn what to expect in the second year of AM. During mentees' second and final year of the program, mentoring sessions are intended to resume as soon as the school year begins.

IMPACT STUDY OVERVIEW

The impact study is a student-level RCT designed to estimate the impact of AM on student-level educational outcomes (school attendance, credits earned, and discipline) and theoretically relevant mediators (e.g., perceived support, academic dispositions). We employ regression analysis to estimate the impact of the program; covariates, including the baseline measure of the outcome variable, and randomization blocking variables were included in the analytic model to increase the precision of our estimates. We estimate program impact within an intent-to-treat (ITT) framework where students are analyzed in their randomly assigned study condition (AM treatment or class-as-usual control), regardless of their actual exposure to the AM treatment.

Within each participating study school, 10th-grade students who were eligible and consented were individually randomly assigned to be offered either the treatment (AM) or control condition (class as usual). Randomization was blocked by school and cohort year at an assignment ratio of 1:1. Students were recruited from 13 public high schools across 5 states on the East Coast (see Figure 1). Students assigned to the treatment group were offered AM during their first and second years enrolled in the study, when they were enrolled in 10th and 11th grades.[3] To assess impact on educational outcomes (attendance, credits, discipline), student educational and demographic data were requested from each participating school or school district. To assess impact on social and emotional outcomes (perceived support, academic dispositions), we administered self-report questionnaires to study participants up to three times – at the beginning of their 10th-grade year (pre-program), at the end of 10th grade, and again at the end of 11th grade. School-level fidelity and individual-level dosage data were also collected to assess the extent study participants were exposed to the intervention.

Figure 1. AM Study Region



RESEARCH QUESTIONS

The impact evaluation answers several research questions concerned with AM's effect on outcomes identified by the program's theory of change and logic model, presented in Appendix A. The AM program implemented under this project is intended to be a two-year program, where students who are

[3] Throughout this report we refer to the study observation period as being students' 10th- and 11th-grade years for simplicity. In actuality, a student may be held back at the end of a given year or promoted midyear, depending on the school's policies and the student's performance and therefore students may be designated as something other than 10th or 11th grade during their first or second year of study participation, respectively.

identified as meeting one of three risk indicators in ninth grade (high absenteeism, course failure, and/or repeated disciplinary infractions) are provided individualized mentoring and support for two full academic years (10th and 11th grades).

Our original evaluation design, which was preregistered on the Registry of Efficacy and Effectiveness Studies (REES), outlined our plan to enroll two study cohorts totaling 800 participants and assess outcomes after two academic years. Some study schools, however, decided to withdraw their participation in the project and only implemented the program for one year. Additionally, after achieving a smaller sample size than intended after the first two cohorts, PRG and CSS decided to enroll a third cohort of schools and students that could only be assessed for one year within the grant's timeframe. As a result, our research questions assess the program's impact on educational outcomes (attendance, credits, and discipline) and social and emotional outcomes (perceived support, academic dispositions, etc.) at two different time points, at the end of the first and second academic years after enrollment (10th and 11th grade, respectively), to better understand the program's pathway of change.

The research questions are as follows:

1. What is the impact of the offer to participate in AM (treatment) relative to the offer to participate in the control (class as usual) on participants' ***school attendance*** at the end of 10th and 11th grades?

2. What is the impact of the offer to participate in AM (treatment) relative to the offer to participate in the control (class as usual) on participants' ***progressing in school*** at the end of 10th and 11th grades?

3. What is the impact of the offer to participate in AM (treatment) relative to the offer to participate in the control (class as usual) on participants' ***student discipline*** at the end of 10th and 11th grades?

4. What is the impact of the offer to participate in AM (treatment) relative to the offer to participate in the control condition (business as usual) on the following ***social and emotional outcomes*** targeted by AM and outlined in the logic model at the end of 10th and 11th grades?
    a. Perceived support from adult(s)
    b. Perceived peer acceptance
    c. Academic self-concept
    d. Academic self-efficacy
    e. Self-efficacy in help-seeking
    f. Self-efficacy in goal-setting
    g. Decision-making skills

We assess each of these four research questions at two time points, which have different sample compositions. The 10th-grade sample includes all students who were randomized into the study and who were observed (and have data) for one full academic year after their enrollment. This represents the full sample of students enrolled across the three cohorts. The 11th-grade sample includes students who were randomized into the study and who were observed for two full academic years after their enrollment. Because some schools dropped out of the study after their first year of participation, and

some schools joined the study late (in Cohort 3), this group represents a subsample of the full population.[4, 5]

CONTROL EXPERIENCE – CLASS AS USUAL

This study is an individual-level RCT with a class-as-usual control condition. Students assigned to the control condition attended the same schools and enrolled in the same classes as students assigned to the treatment condition. The class-as-usual experience (the specific courses or programming offered to control participants) was realized differently at each school in the study, depending on which class period mentors elected to pull mentees to meet. Students who were assigned to the control condition remained in their usual class(es) during the time when AM students were pulled from class to meet with their mentors for mentoring sessions. The flexibility in program implementation regarding what classes and when students are pulled to participate is a typical expression of AM implementation; however, CSS recommends that students only be pulled from nongraded courses or activity periods.

One noteworthy aspect of this experience is that students in the control condition would have received more time (approximately 20 minutes per week) in their scheduled classes or activities over the course of the school year than the treatment students. It is possible this variation could have influenced outcomes by increasing the effort required by the student to achieve academically (since they did not have as much built-in study time) or by reducing the amount of enjoyment and engagement that a student has with school. This is not a design confound but rather an artifact of the AM program itself because it results from the intended implementation of the program.

IMPACT STUDY DESIGN

The impact study investigates whether offering AM to students at risk of dropping out improves their attendance at school, credits earned, suspension rate, and social and emotional outcomes during their 10th- and 11th-grade years. We do this by comparing outcomes for students randomly assigned to be offered the AM program (treatment) with those of students assigned to be offered class as usual (control). The study is an individual-level RCT in which students are the unit of randomization and analysis.

We assess impact within an ITT framework to measure the effect of the offer of the treatment condition relative to the offer of the control condition (as opposed to the effect of exposure to the assigned condition). Outcome data were collected from the study school or district data managers who provided individual-level student administrative records and self-reported from participants using the *Participant Questionnaire*. Data collection procedures were the same for students enrolled in both the treatment and control conditions. Estimates of program impact on each outcome were produced by way of regression equations that model the outcome as a function of treatment assignment, randomization blocking variables, and covariates, including the baseline measure of each outcome.

[4] As documented in REES, we initially designated three confirmatory research questions that examined AM's impact on attendance, progressing in school, and discipline at the end of two years, to assess the efficacy of the full intended program on educational outcomes. Research questions that assess impact at the earlier time point (at the end of one year of programming) and on social and emotional outcomes were designated as exploratory.

[5] See footnote 3.

ELIGIBILITY CRITERIA

The AM program is intended to be implemented with students who are identified as being at risk for dropping out of school based on data from their first year of high school. CSS, the grantee and programmer, was responsible for selection, recruitment, confirmation, and retention of study high schools. They identified 13 schools in Delaware, Maryland, New Jersey, Pennsylvania, and rural North Carolina to participate in this study across three cohorts. CSS actively recruited schools serving large numbers of students representing subpopulations at disproportionate risk for poor academic outcomes, including exclusionary discipline practices (i.e., suspensions and expulsions). For each school that expressed interest in project participation, CSS conducted a readiness assessment prior to confirmation. Once confirmed, they provided each study school with a fully developed curriculum, staff training, technical assistance, and financial support for up to two years that the school implemented AM and participated in the study.

PRG coordinated with school staff to screen all 10th graders at the start of the fall semester to determine whether or not each student on the school's 10th-grade roster met the program and study eligibility criteria. PRG created an *Eligibility Screening Tool* to house each school's student-level eligibility data based on these criteria. Study schools were provided with two options for screening students. To complete the screening process, study schools could either: (1) provide relevant administrative data to PRG, whereby PRG completed the *Eligibility Screening Tool*; or (2) fill out the *Eligibility Screening Tool* using administrative data and submit the completed tool to PRG. In all cases, schools submitted eligibility information through a secure file sharing system managed by PRG.

To be eligible for enrollment in the study, students had to meet all five of the following criteria:

1. Be enrolled in 10th grade at a study school at the time of randomization
2. Not already be enrolled in the study[6]
3. Be considered at risk for dropping out of high school by meeting one or more of the following performance, attendance, and/or disciplinary criteria: [7]
    a. Failed one or more core courses in the previous school year
    b. Missed more than 20 days in the previous school year, but attended school 3 days per week on average
    c. Incurred three or more disciplinary infractions (discipline referrals, suspensions, or detentions) in the previous year
4. Provide consent (either passive or active) to participate in the study
5. Not be simultaneously enrolled in a similar youth development program developed by CSS (i.e., Peer Group Connection-High School)

We present additional details on each of these inclusion criteria in Appendix C.

ASSIGNMENT PROCEDURES

At the beginning of the 2021–22, 2022–23, and 2023–24 school years, students in the 10th grade who met all eligibility criteria were individually randomly assigned to either the AM (treatment) or class-as-usual (control) condition. Randomization was blocked by school and school year. If a student met all eligibility criteria, PRG assigned them a unique study ID number and randomly allocated them to either

[6] This criterion primarily applied when a school contributed more than one cohort of students to the study and if a student was held back to repeat 10th grade for a second year.

[7] These are program eligibility criteria as well.

the treatment or control condition at a 1:1 ratio where each student had a 50% chance of being assigned to either condition.

In some cases, a school could not offer a spot in the study to all eligible students. Each study school determined its maximum number of mentee (treatment) spots according to staff mentoring capacity. The number of study spots available at a school was determined as twice the number of mentee spots (to maintain a 1:1 assignment ratio). If the number of eligible students at a school exceeded this maximum number of study spots, PRG conducted a two-stage randomization procedure. In this scenario, we first randomly selected a set of students to participate in the study based on the number of available study spots (from the full eligible roster), and then randomly allocated this set of participants into either the treatment or control group (at a 1:1 assignment ratio). Students were considered enrolled in the ITT sample at the point of random assignment into either the treatment or control condition.[8] Random assignment was conducted using the *ralloc* command in Stata.

After randomization, PRG provided CSS and school staff a roster of study participants at that school and their assignments. Prior to the start of AM mentoring sessions, CSS and PRG coordinated with school staff to confirm that study participants' mentor assignment accurately reflected individual random assignment; any mistakes in assignment were corrected as quickly as possible. In addition, on an ongoing basis, PRG and CSS collected and reviewed AM mentoring session logs in order to track dosage and ensure there was no crossover. In instances where crossover was detected or participation was low, PRG and CSS worked with study schools to correct the problem if possible.

OUTCOME MEASURES

This impact study examines the effect of offering AM on students' educational, behavioral, and social and emotional outcomes at two different time points (end of 10th and 11th grades). Outcomes are operationalized as follows: (1) school attendance – a count variable that indicates the number of days a student is present at school during each school year a student is enrolled in the study; (2) progressing in school – a count variable that indicates the number of credits earned during each school year; (3) student discipline – a dichotomous variable that indicates whether or not the student was suspended during each school year; and (4) social and emotional outcomes – a set of seven measures that examine students' attitudes and perceptions related to academic learning and perceived support and acceptance at their school. Outcome data were collected uniformly either from data managers or using self-report questionnaires for all study participants.[9]

SCHOOL ATTENDANCE

Student attendance is operationally defined as the number of days a student is recorded as being present at a school during a given school year. Schools are responsible for tracking attendance for all students enrolled at their school and reporting it to their state education agency. Regardless of state or school district, all students under the age of 16 are considered continuously enrolled until the last day of the school year. However, the definition of *days present* may vary by state and individual school

[8] The exceptions were students who were found to have withdrawn from the school prior to the date of randomization. In the event a school or district data manager confirmed that a student withdrew from the study school prior to the date of randomization, that student was not considered enrolled in the study and does not count toward our attrition calculations.

[9] For additional details of variable operationalization, see Table C.2 in Appendix C.

districts.[10] Although the definition may vary across states and school districts, there is no variation within schools (attendance is defined the same way for treatment and control students within each school).[11]

At the first time point (end of 10th grade), days of attendance is equal to the total number of days the student attended school during their first year of the study. At the second time point (end of 11th grade), days of attendance for each of the two academic years were summed to construct a continuous measure of the number of days a student attended school during both study years. Students for whom a complete measure of attendance could not be constructed due to incomplete or missing attendance data (due to transferring out of the study school) counted toward attrition for the school attendance outcome. Students who withdrew from school and did not transfer to a new school (i.e., dropped out) were included in the analytic sample if the school provided data for the number of days the student attended prior to dropping out of school.

CREDITS EARNED

Progressing in school is operationally defined as the number of credits earned toward high school graduation during a given school year. PRG worked with study schools to collect the number of credits earned during each school year. All credits earned during the school year were counted in the construction of the outcome measure, regardless of whether a student was promoted to the next grade at the end of the year.

At the first time point (end of 10th grade), credits earned is equal to the number of credits earned during their first year of the study. At the second time point (end of 11th grade), credits earned during the first and second academic years were summed to construct a continuous measure of the number of credits earned during both years. Similar to attendance, students for whom a complete measure of credits could not be constructed due to incomplete or missing data (due to transferring out of the study school) counted toward attrition for the progressing in school outcome. Students who withdrew from school and did not transfer to a new school (i.e., dropped out) were included in the analytic sample if the school provided data for the number of credits earned prior to dropping out of school.

SUSPENSION

Student discipline is operationally defined as a dichotomous variable that equals zero if the student incurred no suspensions during a given school year, and one if the student was suspended for one or more days during a given year. Schools tracked suspensions (in-school and out-of-school) while students were enrolled in school. Although the definition of suspension may vary across states and school districts, there is no variation within schools (days suspended are defined the same way for treatment

[10] As an example, in North Carolina, students are considered to have attended a school day if they are present for at least 50% of the school day whereas in New Jersey, a student is considered to have attended a school day if they attend school for at least one hour in the morning and one hour in the afternoon. Information about North Carolina student accounting data can be found in the *School Attendance and Student Accounting Manual*. Retrieved April 16, 2024, from https://www.dpi.nc.gov/districts-schools/district-operations/financial-and-business-services/student-accounting#SASAManual-1394. Delaware's school attendance policy can be found in *The Delaware Code Online*. Retrieved May 1, 2024, from https://delcode.delaware.gov/title14/c027/sc01/. Information about Maryland student accounting data can be found in the *Maryland Student Records System Manual*. Retrieved April 16, 2024, from https://marylandpublicschools.org/about/Pages/OCP/Publications/index.aspx. Information about New Jersey student accounting data can be found in the *SID Management Student Data Handbook.* Retrieved April 16, 2024, from https://www.njsmart.org/njr/ks/SID%20Management/NJ%20SMART%20SID%20Management%20Student%20Data%20Handbook.pdf. The School District of Philadelphia attendance policy can be found in its *Code of Conduct*. Retrieved May 1, 2024, from https://www.philasd.org/schoolboard/wp-content/uploads/sites/892/2022/09/Code-of-Conduct-2022-2023.pdf.

[11] For additional details on operationalization of the school attendance outcome, see Table C.2 in Appendix C.

and control students within each school).[12] At the first time point (end of 10th grade), we examined the number of days a student was suspended during their first academic year. Students were coded as 0 if they incurred zero days during the first school year and coded as one if they incurred at least one suspension during the year. At the second time point (end of 11th grade), we examined the number of days a student was reported to be suspended during the first and second academic years. Students were coded as 0 if they incurred zero days during both years and coded as 1 if they incurred at least one suspension during either year.

Unlike attendance and credits, if a student dropped out of school during their 10th- or 11th-grade year, the student was considered as having incomplete discipline outcome data because they did not have the "opportunity" to incur suspensions during the period that they were not enrolled in school. Including these students in the analytic sample could have skewed findings if members of one treatment group had dropped out of school at a higher rate than the other and therefore could have fewer days suspended. As a result, we count these students with incomplete data as being attrited from the analytic sample.

SOCIAL AND EMOTIONAL OUTCOMES

We also examine seven social and emotional outcomes, each of which measures a hypothesized mediator of educational behaviors or performance targeted by AM, gathered via self-report questionnaires. Outcome measures and their definitions are listed in Table 1.

Table 1. Social and Emotional Outcome Definitions

Outcome	Domain[13]	Definition
Perception of support from adults at school	School Climate	Students' perceptions of the social support they receive through at least one relationship with a teacher and/or other adult at school.
Perception of peer acceptance	School Climate	Students' perceptions of their peer group in school and the degree to which they feel accepted by those peers.
Academic self-concept	Academic Dispositions	Students' set of self-thoughts and beliefs about their overall confidence, effort, interest, and identity related to education.
Academic self-efficacy	Academic Dispositions	Students' feelings of competence in their academic ability and feeling as though they are capable of learning required skills.
Self-efficacy in help-seeking skills	Academic Dispositions	Students' perceptions of their own ability to ask for help, including knowing who or where to ask for help and their level of comfort in asking for help.
Self-efficacy in goal-setting skills	Academic Dispositions	Students' perceptions of their own ability to set and work toward goals, including the frequency with which they set them.
Decision-making skills	Academic Dispositions	Students' evaluation of their ability to think about options available to them and their potential outcomes prior to making a decision and how often students employ the decision-making skills they have acquired.

[12] For information on each state's disciplinary accounting data, see sources cited in footnote 10.

[13] Outcome domains are based on version 5.1 of the WWC's Study Review Protocol (December 2024). This protocol defines school climate as, "observations or assessments of the schoolwide or postsecondary institution environment or culture, as distinct from one's own behavior, such as the quality of social interactions, attendance, safety, engagement in school, sense of belonging, staff cohesion, teacher-student relationships, and parent-teacher communication, as well as the prevalence of bullying at the school" (pg. 7). The protocol defines academic dispositions as, "indicators that are focused on self-reported or assessed student attitudes toward academics or participation in school activities. Outcomes in

All of the social and emotional outcomes were operationalized as mean scale scores from questionnaire items with 7-point Likert-type scales. Scale scores were constructed by estimating the mean of all items that made up the scale and were only estimated if a student responded to all items within a specified scale. We did not impute any missing values in these or any outcome measures.[14]

DATA SOURCES AND COLLECTION

Data were collected from two sources: (1) student baseline and outcome administrative records compiled by study school and/or district data managers; and (2) self-report participant questionnaires completed by study participants. Prior to each school's participation in the study, CSS obtained a Memorandum of Understanding (MOU) with the school and PRG established a study agreement or data sharing agreement with the school or district. Study agreements specified details pertaining to the study timeline, AM mentor to mentee capacity, eligibility determination, randomization of students, and data collection (survey and student records). Methods of data collection were identical for treatment and control students. Figure 2 presents the data collection time points as they relate to program implementation and school years for each study cohort.

Figure 2. Multiyear Intervention Study Design

		SY 2020–21	SY 2021–22	SY 2022–23	SY 2023–24
Cohort 1	Intervention		1st year of implementation	2nd year of implementation	
	Data collection	9th-grade outcomes (baseline)	10th-grade outcomes	11th-grade outcomes	
Cohort 2	Intervention			1st year of implementation	2nd year of implementation
	Data collection		9th-grade outcomes (baseline)	10th-grade outcomes	11th-grade outcomes
Cohort 3	Intervention				1st year implementation
	Data collection			9th-grade outcomes (baseline)	10th-grade outcomes

SCHOOL RECORDS

Student administrative records were prepared by school- or district-level data managers and sent to PRG following secure data sharing protocols. Administrative records contained background demographic data, indicators of academic disadvantage (Individualized Education Plan [IEP] status, English Language Learner [ELL] status), academic and administrative data (enrollment, attendance, credits earned), and behavioral outcome data (suspensions). During the fall semester of each cohort (2021, 2022, and 2023), PRG submitted a baseline administrative data request to each data manager to collect demographic characteristics and prior year (ninth grade) outcome data for each student enrolled in the study. PRG sent a second data request to each data manager during the summer following participants' 10th-grade year and, if applicable, a third request during the summer following participants' 11th-grade year.

this domain include academic growth mindset, academic motivation, academic or subject-specific self-efficacy, academic engagement, and academic grit" (pg. 6).

[14] For additional details on the construction of these noncognitive outcomes, see Appendix C.

PARTICIPANT OUTCOME QUESTIONNAIRE

PRG developed and administered a *Participant Questionnaire* to collect data related to social and emotional outcomes, as well as to gather information about barriers to school engagement and demographic characteristics. Study participants completed the self-report instrument up to three times – once at the beginning of the fall semester of their 10th-grade year (pre-program), again at the end of 10th grade, and, if applicable, a third time at the end of their 11th-grade year. At each time point, PRG trained at least two staff at each study school to act as proctors and administer the questionnaire to study participants. PRG selected proctors that were not staff involved in the implementation of AM (i.e., were not a mentor or on the stakeholder team) and ideally were not 10th- or 11th-grade teachers with whom participants would have frequent interaction. Proctors administered the questionnaire with study participants in either a group (classroom) environment or individually, depending on student schedules. Students completed the questionnaire electronically in Qualtrics, a web-based survey software, and completed questionnaires were sent directly to PRG.

For students who were no longer enrolled in the study school or who were absent during the window of time school staff were administering the questionnaires, PRG staff attempted to reach the student individually through email and/or phone contact information provided on the student's *Locator Form*, completed at baseline. These students were sent a personalized link to complete the questionnaire so that study staff could track completions. Students were provided with a $10 gift card each time they completed a questionnaire as a thank you for study participation.

ANALYTIC METHODS

The impact study aims to determine whether offering AM to students at the beginning of their 10th-grade year impacts the number of days the student attends school, the number of credits earned, whether or not the student was suspended, and mean scores on social and emotional outcome measures. We estimate program impacts within the ITT framework, which means that all students who have been randomized and provide outcome data are included in the analysis according to the group they were assigned regardless of their actual exposure to the AM program. An ITT estimate is preferred because it minimizes the potentially biased post-enrollment self-selection that motivates some students to engage more and others to engage less with the intervention.

We estimate the impact of the AM program by regressing each outcome on treatment assignment, the baseline measure of the outcome variable, demographic covariates (age, race/ethnicity, gender), academic disadvantage indicators (IEP and ELL status), and randomization blocking variables. Although a straight difference-of-means approach generally provides unbiased estimates of the effect of the treatment intervention within an RCT, a statistical model that includes covariates and the baseline outcome is preferred because it may increase the precision of the impact estimates. An ordinary least squares (OLS) model was used to estimate the impact of the program on all outcomes (using Stata 18).[15] If tests are statistically significant at the 0.05 level, then we consider the program to have had an impact on the outcome. Given we did not achieve our target sample size of 800, we also consider marginally significant results ($p < 0.10$) to indicate promising impact on an outcome. Along with significance tests, we examine magnitude of effect (effect size) to better understand the practical significance of the findings. For additional details on the analytic methods, including the model specification, see Appendix C.

[15] We run a series of sensitivity analyses on each outcome to examine variations in impact estimates based on different modeling approaches, including more parsimonious models with fewer covariates included to test the robustness of our benchmark approach. In addition, we also conduct a sensitivity test where we use a logistic regression model on the dichotomous outcome of suspension rate.

STUDY PARTICIPANTS

We conducted the same random assignment process in fall 2021 (Cohort 1), 2022 (Cohort 2), and 2023 (Cohort 3). Table 2 presents the total number of students who were randomized into the study at each school during each of the three cohorts, overall and by treatment condition.

Table 2. Study Enrollment

School	State	Urbanicity	AM	Control	Total
Cohort 1 (SY 2021–22)					
School A	NC	Rural	4	4	8
School B	NC	Rural	6	6	12
School C	NC	Rural	17	18	35
School D	NC	Rural	10	9	19
School E	DE	Urban	13	13	26
Cohort 2 (SY 2022–23)					
School A	NC	Rural	7	7	14
School B	NC	Rural	3	3	6
School E[16]	DE	Urban	7	7	14
School F	NC	Rural	20	20	40
School G	NC	Rural	20	20	40
School H	NC	Rural	20	21	41
School I	PA	Urban	12	11	23
School J[17]	PA	Urban	4	4	8
School K	NJ	Urban	22	21	43
Cohort 3 (SY 2023–24)					
School B	NC	Rural	9	9	18
School C	NC	Rural	7	7	14
School L	MD	Suburban	9	8	17
School M	MD	Suburban	7	8	15
Total	*N/A*	*N/A*	*197*	*196*	*393*

[16] School E communicated to the project team that they intended to participate for a second year of the study and enroll a new cohort of 10th graders at the start of the 2022–23 school year. The principal submitted the results of the eligibility screening to PRG and we subsequently conducted random assignment of eligible students. However, the school administrators discontinued correspondence with CSS and PRG immediately following randomization procedures. It is our understanding that no program activities occurred after the school became uncommunicative; however, we retained the 14 students randomized at the start of the 2022–23 school year in our ITT sample and include them in our calculations of attrition. We do not, however, include this school (during Cohort 2) in our discussion of program implementation and fidelity going forward.

[17] School J also intended to participate as a new school during Cohort 2 and submitted eligibility screening data for PRG to randomly assign eligible and consented students. However, the STC did not subsequently match treatment students with a mentor and no programming occurred during the 2022–23 or 2023–24 school year. As a result, we retain these eight students in our ITT sample and will include them in our calculations of attrition. We do not include this school in our discussion of program implementation and fidelity going forward.

As shown in Table 2, in total, 393 tenth graders were enrolled in the impact study; 197 students were assigned to the treatment condition and 196 were assigned to the control condition. Of the 13 study schools, 9 participated for a single cohort (i.e., contributed one block of students) and 4 participated for more than one cohort (contributed multiple blocks of students). Out of the 393 students enrolled in the study, 63% were enrolled in rural schools, whereas the remaining 37% attended schools in urban/suburban communities.

Table 3 provides descriptive characteristics and baseline outcome data of the full set of study participants who were randomized across all three cohorts. As shown in Table 3, the pooled sample of randomized participants included 393 tenth-grade students. About half identified as Black (52%) whereas a quarter identified as White (22%) or Hispanic/Latino/a (23%). A small proportion were designated as English learners or had an IEP (10% each). In terms of their educational outcomes during ninth grade, the average attendance rate was 85% whereas the suspension rate was 19%; students earned just under nine credits in ninth grade, on average.[18]

Students could be eligible for the program by meeting at least one of three eligibility criteria. About two thirds (65%) of our study sample met the criteria of failing a course in ninth grade, whereas about half missed 20 or more days of school (55%).[19] A smaller portion (20%) met the criteria of incurring three or more disciplinary infractions during ninth grade. Students were asked to self-report any barriers they felt make it difficult to attend school and complete schoolwork on their baseline questionnaires at the beginning of 10th grade. Students were most likely to report feeling stressed or anxious (42%) or depressed (31%) as barriers to learning. Other reported barriers included students' own health or the health of a close family member (30%), having responsibilities outside of school (30%), or feeling like they have little control over their life (26%).

[18] The number of credits required for graduation varied across states represented in this study. For example, in North Carolina, students need 22 credits to graduate from high school, whereas in New Jersey, students need 130 credits (with one course generally equating to 5 credits).

[19] The typical school year in the districts that participated in this study ranged from 163 to180 days.

Table 3. Descriptive Characteristics of Study Participants

Characteristic	Number Reporting	Statistic
Age		
Mean age in years (at baseline)	393	16.14
Race/ethnicity[20]		
Black or African American	204	51.9%
Hispanic or Latino/a	92	23.4%
White	89	22.7%
Other race[21]	31	7.9%
Multiracial	18	4.6%
Gender		
Male	204	51.9%
Academic disadvantage		
English learner	38	9.7%
IEP	41	10.4%
9th-grade outcomes		
Attendance rate	393	85.0%
Mean credits earned	393	8.76
Suspended	74	18.8%
Self-reported barriers to school		
Own or family member's health	117	29.8%
Feeling down or depressed	120	30.5%
Feeling stressed or anxious	167	42.5%
Feelings of loss of control over life	102	26.0%
Responsibilities outside of school	118	30.0%
Limited access to technology at home	15	3.8%
Don't feel safe at school	13	3.3%
Don't feel safe in community	12	3.1%
Don't feel safe at home	10	2.5%
Distractions at home	67	17.1%
No adult at home to help with school	19	4.8%
Other	39	9.9%
Average number of barriers selected	*393*	*2.31*
AM eligibility criteria met (in SY prior)		
Missed 20+ days of school	217	55.2%
Failed a core course	254	64.6%
Had 3+ disciplinary infractions	79	20.1%

IMPLEMENTATION STUDY RESULTS

In this section, we present findings from the implementation evaluation of AM, including details of program structure at each study school, adherence to the AM program model, and program dosage.

In all, 13 schools agreed to participate in the study, signed an MOU with CSS, and recruited and enrolled students into the study. Of these schools, 12 implemented at least some programming. The program was intended to be implemented at each school for two academic years; however, some did not maintain participation in the project for both years. The following schools withdrew from the study early:

- School E enrolled students in the study for both Cohorts 1 and 2 but withdrew their participation prior to the beginning of programming in the second year of the study (prior to the beginning of the second year of implementation for Cohort 1 and the first year of implementation for Cohort 2).
- School J enrolled students in the study for Cohort 2, but withdrew prior to any implementation that year.
- School G enrolled students in the study in Cohort 2 and implemented the first year of programming but withdrew from the study prior to the second year of programming.

All students who were enrolled in the study are retained in the ITT sample, regardless of their actual participation in the program and collection of outcome data. In the impact study, students who do not have data are considered lost to attrition. In the Program Implementation and Adherence to Achievement Mentoring Model sections that follow, we only present information on schools during the years they actively participated in the study. The section on Program Dosage examines the amount of programming received by the ITT sample to provide important context for the impact study results.

PROGRAM IMPLEMENTATION

Figure 3 presents, for each cohort, each participating school's period of active programming during each year of implementation. Table 4 then provides a brief overview of the number of mentors and mentees active at each school and the average number of mentoring sessions held during each implementation year. Note that schools E (Cohort 1) and G (Cohort 2) discontinued implementing the program after their first year and did not participate in the study for a second year. For Cohort 3 schools, the second year of implementation (scheduled to occur in 2024–25) is outside the observation period of the grant, which ended in September 2024.

[20] Districts reported student race and ethnicity in the same field on the data request form. As a result, categories are not mutually exclusive and totals exceed 100%. A total of 47 students (12.0%) did not identify their race.

[21] *Other* includes students who identified as Asian, American Indian/Alaska Native, or Pacific Islander.

Figure 3. *Implementation Periods*

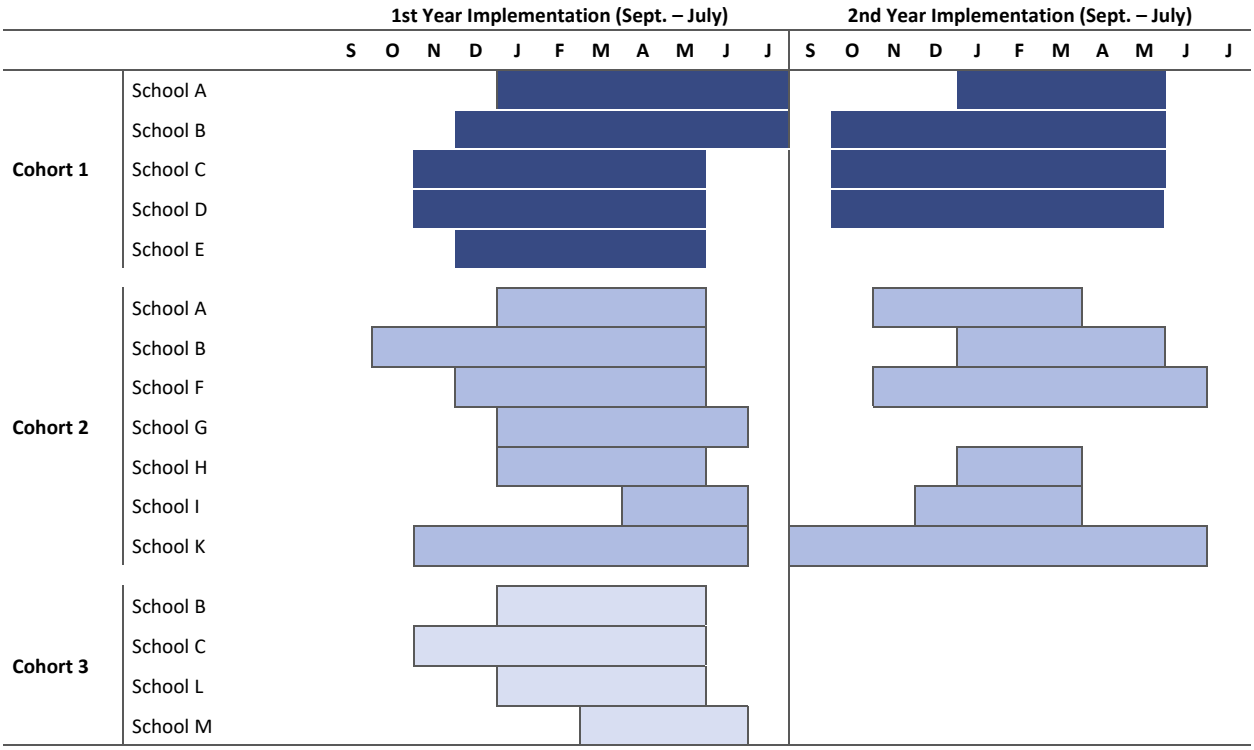


Table 4. *Implementation Details*

		Year 1			Year 2		
	School	Mentors	Mentees	Average Sessions (SD)	Mentors	Mentees	Average Sessions (SD)
Cohort 1	School A	3	3	13.7 (7.6)	2	2	17.0 (0.0)
	School B	4	6	13.2 (7.1)	3	6	24.8 (10.8)
	School C	8	17	14.0 (4.3)	6	15	20.5 (6.5)
	School D	8	9	9.8 (6.9)	5	6	29.5 (5.5)
	School E	7	12	9.6 (4.2)	–	–	–
Cohort 2	School A	5	7	17.4 (1.1)	4	6	9.3 (9.3)
	School B	2	3	28.0 (9.2)	1	2	15.5 (0.7)
	School F	11	20	14.9 (5.8)	10	18	18.8 (12.5)
	School G	9	18	10.2 (6.6)	–	–	–
	School H	9	19	14.6 (7.0)	9	16	20.8 (10.4)
	School I	8	11	5.8 (3.8)	7	8	11.5 (7.7)
	School K	10	20	9.4 (6.4)	8	13	11.1 (8.7)
Cohort 3	School B	4	8	11.3 (4.7)	–	–	–
	School C	6	6	18.8 (2.0)	–	–	–
	School L	5	7	10.3 (6.6)	–	–	–
	School M	3	7	11.9 (0.4)	–	–	–
	Total	*6.4*	*10.8*	*12.4 (6.6)*	*5.5*	*9.2*	*18.1 (10.5)*

Data presented in Figure 3 and Table 4 indicate that schools varied in their overall length of programming and capacity to reach mentees. Three schools participated in more than one cohort, resulting in a total of 16 unique cohort-school blocks that implemented programming. Of these 16 blocks, 10 (62.5%) implemented AM for two years as intended, and 6 (37.5%) implemented the program for a single year. For the four Cohort 3 sites, this was a known implementation limitation upon enrollment at the start of 2023–24 as the second year of programming (scheduled for 2024–25) was outside the scope of the grant period, which ended in September 2024. Two sites that had intended to implement AM for a second year (Schools E and G) severed partnership with CSS and PRG after the first year of participation in the program and study. In one instance, the administrator who was also the STC at the school stopped responding to communication attempts in early fall of the second year and in the second instance, the school district overrode the school's desire to participate in the project and revoked their MOU with CSS.

Across cohorts and implementation years, schools had an average of 6 mentors assigned to 10 mentees. This ranged widely based on school capacity and the number of eligible students (e.g., met criteria and consented to participate). Schools implemented the program for an average of five to six months during the school year (this ranged from a minimum of two months to a maximum of eight). Program start and end dates are based on the timing of the school's kickoff and closure events each year. Mentors held on average between 14 and 15 mentoring sessions with their mentees over the course of a school year. This ranged widely across schools and implementation years with a minimum of 5.8 at one school during their first year and a maximum of 29.5 at another school during their second implementation year. The average number of mentoring sessions held increased, on average, between the first year of implementation (12.4 sessions) and the second year (18.1 sessions). In a few cases, schools delayed program kickoff until after the winter break and mentors began meeting with their mentees at the start of the spring semester.

ADHERENCE TO ACHIEVEMENT MENTORING MODEL

Table 5 presents an overview of adherence to the AM program model, including the percentage of sessions where mentors completed core activities outlined in the AM handbook each school year. Session activities are ordered in the sequence in which they are intended to be carried out. Mentors are advised to complete a series of steps during each mentoring session, including interviewing one of their mentee's teachers beforehand to get feedback on the student's classroom and academic behaviors, reviewing that feedback with their mentee, giving the student praise for instances of school engagement from the past week, helping the student to set a SMART (specific, measurable, attainable, relevant, and time-bound) goal for the week ahead, and to create a plan for achieving that goal. Mentors are also expected to reach out to their mentee's parents/guardians approximately once per month to share positive feedback about their child.

Table 5. Session Adherence

Activity	SY 2021–22	SY 2022–23	SY 2023–24	Overall
Percent of sessions where mentor completed all activities	58.6%	65.5%	46.0%	57.4%
Interviewed teacher before session	77.1%	79.5%	68.5%	75.2%
Shared *Weekly Report Form* (WRF)	69.9%	76.2%	59.1%	69.1%
Gave praise	95.2%	97.6%	90.2%	94.6%
Reviewed teacher comments	67.0%	71.9%	56.8%	65.8%
Student responds to WRF	67.8%	74.0%	53.9%	65.9%
Helped student set a SMART goal	92.1%	92.3%	75.5%	86.4%
Created implementation plan	91.5%	92.3%	75.4%	86.3%
Average number of parent contacts	4.5	5.0	4.7	4.8

As presented in Table 5, mentors documented varying degrees of adherence to the AM model during each session held with their mentee. In terms of the full checklist of tasks, mentors indicated that they adhered to the AM session structure for more than half the sessions held (57.4%). Mentors were most likely to give their mentee praise for instances of engagement with school during each session (94.6%) followed by helping their mentee set a goal for the coming week (86.4%) and establish a plan to achieve that goal (86.3%). Although mentors completed all other activities in at least two thirds of sessions, they were slightly less consistent with interviewing teachers prior to the session (75.1%), sharing teacher feedback from interviews (65.8%), sharing the WRF (69.1%), and eliciting student feedback from the WRF (65.9%). Adherence to the model varied slightly across each school year during the project, with adherence peaking during the 2022–23 school year. Finally, mentors reported contacting their mentee's parents/guardians approximately 4 to 5 times during the school year with peak adherence occurring during the 2022–23 school year.

PROGRAM DOSAGE

Table 6 presents a summary of program dosage for the ITT sample of participants who were randomly assigned to receive AM (treatment), within each cohort and overall. We present, for each cohort and implementation year, the percentage of treatment participants who received any programming and the recommended minimum dosage, as well as the average number of sessions attended and the percentage who participated for both years. Note for Cohort 3 schools, the second year of implementation (scheduled to occur in 2024–25) is outside the observation period of the grant, which ended in September 2024.

Table 6. *Program Dosage*[22]

	Cohort 1	Cohort 2	Cohort 3	Overall[23]
Number of students randomly assigned to AM	50	115	32	197
Percent who received any mentoring	92.0%	78.3%	78.1%	81.7%
Year 1				
Average number of sessions	11.2	10.1	10.0	10.3
Percent who attended at least 20 sessions	4.0%	12.2%	9.4%	9.6%
Year 2				
Average number of sessions	13.4	8.3	–	9.8
Percent who attended at least 20 sessions	36.0%	21.7%	–	26.1%
Percent who participated for two years[24]	56.0%	44.4%	–	47.9%

Individual-level dosage data suggest that students assigned to the treatment condition did not receive the recommended minimum threshold dosage of mentoring (20 sessions per year). As shown in Table 6, a total of 197 students were randomly assigned to be offered AM at the beginning of their 10th-grade year across all three cohorts. Of these, 161 (81.7%) received some level of mentoring after random assignment, whereas 36 (18.3%) did not receive any mentoring. Reasons for why some students received no mentoring include the student declining participation after being offered the program or leaving the study school (e.g., transferring from or dropping out) shortly after the point of random assignment before the program began. Additionally, 11 treatment students were at schools that did not implement any programming after randomization. On average, students attended about half (10) of the minimum number of recommended mentoring sessions (20) advised by the program developers. Specifically, students attended between 10 and 11 mentoring sessions during their first year of participation and between 9 and 10 sessions in their second year.[25]

Only a small portion of the treatment sample received the recommended amount of mentoring (20 or more sessions) during either year of participation. Out of the 197 students assigned to AM across all cohorts, only 19 (9.6%) met with their mentor 20 or more times during the first year of participation. Out of the 165 students assigned to AM in the first two cohorts, about one fourth (26.1%) met with their mentor 20 or more times during the second year. In total, students assigned to the treatment group attended an average of 18.6 mentoring sessions during the study.[26]

[22] Of the 196 participants assigned to the control condition, 4 (2.0%) were ultimately placed in the AM program and assigned a mentor. These four students were all in Cohort 3 and attended an average of 11.0 sessions during 2023–24.

[23] The Year 2 total column excludes students enrolled in Cohort 3, since there was not enough time in the grant to measure a second year of implementation for this cohort.

[24] The total percentage who participated for two years excludes the 32 students enrolled in Cohort 3 from the denominator. When these students are included in the denominator, the percentage decreases to 40.1%.

[25] One explanation for the decline in average dosage between the first and second year is that only about half (47.6%) of those assigned to AM in the first two cohorts participated for both years. Out of 57 participants in the first two cohorts who received some mentoring, but did not participate for two years, 33 (57.9%) attended one of the two schools that discontinued programming after the first year. The remaining students either transferred or withdrew from the study school after their first year, did not want to participate for a second year, or had their mentor leave the school after the first year.

[26] When we exclude Cohort 3 students, total dosage increases slightly to 20.2 sessions.

MENTOR AND MENTEE PERSPECTIVES

We asked both mentors and mentees to provide insight into their experiences with AM at the end of each school year using brief feedback questionnaires. The feedback collected provided the CSS team with valuable perspective on how AM was impacting both mentors and mentees.[27] On the *Mentor Feedback Form*, mentors were asked to describe the organizational supports in place for AM implementation. Several mentors noted that they had strong collaboration between their administrators and the stakeholder team and/or had time carved out in their day for meetings, which were used to meet with students. Even so, when asked to describe their most significant challenges, many mentors indicated that they struggled to find time to meet with their mentees each week while juggling the demands of their primary role at the school. Some remarked that it was difficult to align their own available time with that of their mentees. Others noted that they were working with students who were chronically absent from school, making in-person mentoring sessions difficult to arrange.

We asked mentors to describe the differences or impacts they noticed in their mentees compared with the start of the school year. Several mentors described how their mentees had become better at communicating and more open to discussing their challenges with school and asking for help. Others noted that they had seen their mentees improve their grades over the course of the school year, attend school more regularly, and/or take school more seriously. More broadly, when asked to describe how they were defining "success" in the AM program, mentors most often defined success as measurable academic growth in terms of bringing grades up and getting on track for graduation. Several also defined success in terms of holistic growth, or seeing their mentees mature as a person, breaking through walls that made it difficult to connect on a personal level, and seeing students take accountability for their actions.

A vast majority of mentees reported that the AM program helped them care more about graduating from high school (85%, n = 177) and that they were doing better at school because of their mentor's help (72%, n = 176).[28] When asked to describe one way that their relationship with their mentor has been important to them, mentees were most likely to describe their positive relationship with their mentor, often describing them as someone they could trust, get along with, and confide in. Mentees valued having a trustworthy adult who listened to them and who they could ask for help when they needed it. Several mentees also noted that their mentor had been important for helping them improve their grades or level of engagement with school, specifically in terms of motivation, confidence, or looking ahead to their future. Mentors largely echoed this same sentiment. When asked what part of AM they felt was most important for students, most indicated the one-on-one adult relationship as the most important aspect, noting that students need an adult who is nonjudgmental, listens without reprimand, and who is reliable. When asked what could have made their relationship with their mentor better, most mentees reported a desire to meet more often with their mentor, indicating the consistency in the relationship is an important factor of the program from mentees' perspectives.

[27] Mentors were emailed a link to the *Mentor Feedback Form* at the end of each school year by their CSS program manager. Responses were sent directly to PRG via Qualtrics. Mentees were automatically routed to the *Mentee Feedback Form* upon completing their *Participant Outcome Questionnaire* in Qualtrics at the end of each school year.

[28] Mentees were asked to rate on a scale from 1 to 5 (*not at all* to *a great amount*) the extent to which AM helped them care more about graduating from high school. We report the proportion of students who indicated AM helped them *quite a bit* or *a great amount*. Mentees were also asked to report how true the statement "I am doing better at school because of my mentor's help" on a scale from 1 to 4 (*not true at all* to *very true*); we report the proportion of students who selected *pretty true* or *very true*.

IMPACT STUDY RESULTS

We present the findings from the impact study in two sections below. For each time point (10th and 11th grades), we first provide a descriptive contrast of the proportion of students enrolled in the ITT sample and those included in the analytic samples for each outcome, including the overall attrition from the randomized sample and the differential attrition between the AM and control groups.[29] We also discuss the baseline equivalence between the AM and control participants included in each analytic sample for the baseline measure of the outcome. We then present the impact estimates for each research question from the Benchmark Analytic Model, followed by a discussion of the key findings.

TENTH GRADE

ATTRITION AND BASELINE EQUIVALENCE

Table 7 presents, for each outcome measure, the number of participants randomized into each treatment condition, the number in each group that had data at the end of 10th grade, as well as the overall and differential attrition rates for the outcome. We note whether the combination of the overall and differential attrition rates was below the *What Works Clearinghouse* (WWC) cautious boundary for an acceptable threat of bias due to attrition, as outlined by the *What Works Clearinghouse Procedures and Standards Handbook, Version 5.0* (WWC, 2022).

Table 7. Randomized and Analytic Samples – 10th Grade

	ITT Sample		Analytic Sample				
	AM	Control	AM	Control	Overall Attrition	Differential Attrition	Under WWC Cautious Boundary
Days attended	197	196	152	157	21.4%	–2.9%	Yes
Credits earned	197	196	152	156	21.6%	–2.4%	Yes
Suspension rate	197	196	148	154	23.2%	–3.4%	Yes
Perceived adult support	197	196	135	128	33.1%	3.2%	Yes
Perceived peer acceptance	197	196	142	140	28.2%	0.7%	Yes
Academic self-concept	197	196	145	137	28.2%	3.7%	Yes
Academic self-efficacy	197	196	145	142	27.0%	1.2%	Yes
Self-efficacy in help-seeking	197	196	144	143	27.0%	0.1%	Yes
Self-efficacy in goal-setting	197	196	145	142	27.0%	1.2%	Yes
Decision-making skills	197	196	142	139	28.5%	1.2%	Yes

As shown in Table 7, across three cohorts of incoming 10th-grade students, we randomized 197 participants into the AM (treatment) condition and 196 into the control, for a total ITT sample of 393 participants. The analytic sample is the subset of students from the original ITT sample for whom we have outcome data and who are retained in the analysis at the end of the 10th grade. The analytic sample varies slightly between outcomes depending on the school's ability to report educational outcome data or item nonresponse on the questionnaires. The overall attrition rates range from 21 to 33%. The differential rates of attrition between the treatment and control groups range from 0 to 4%. Overall attrition rates are slightly higher for self-report outcomes compared with school-reported

[29] Overall attrition refers to the rate of missing data for the entire sample. Differential attrition represents the difference in missing data for the intervention and comparison groups.

administrative outcomes due to students being absent from school during questionnaire administration windows and subsequently unresponsive to follow-up attempts. The combination of overall and differential attrition rates for each outcome are below the WWC's cautious boundary, indicating a low risk of bias due to sample compositional change.

Table 8 presents the baseline balance statistics of each analytic sample, including the sample size, unadjusted mean, and standard deviation of the baseline measure of the outcome, as well as the model-based difference between the treatment and control groups and the standardized difference of means or proportion (i.e., Hedges' *g* or Cox Index).[30]

Table 8. Baseline Equivalence of Treatment and Control Groups – 10th Grade

		Treatment Group			Control Group			Treatment – Control	Standardized
Analytic Sample	**Baseline Measure**	***N***	**Mean**	**SD**	***N***	**Mean**	**SD**	**Difference**	**Difference**
Attendance	Days attended in 9th grade	152	146.47	25.07	157	147.30	26.48	–0.52	–0.02
Credits	Credits earned in 9th grade	152	8.06	7.39	156	8.68	8.57	0.10	0.01
Suspension	Percent suspended in 9th grade	148	17.57%	0.38	154	18.83%	0.39	–0.01	–0.04
Perceived adult support	Baseline mean	135	5.13	1.06	128	5.07	1.21	0.08	0.07
Perceived peer acceptance	Baseline mean	142	4.43	1.35	140	4.40	1.21	0.04	0.03
Academic self-concept	Baseline mean	145	4.59	1.03	137	4.53	1.05	0.04	0.04
Academic self-efficacy	Baseline mean	145	5.55	0.97	142	5.53	1.09	0.04	0.03
Self-efficacy in help-seeking	Baseline mean	144	5.14	1.23	143	5.18	1.30	–0.02	–0.02
Self-efficacy in goal-setting	Baseline mean	145	5.37	1.20	142	5.38	1.40	–0.02	–0.01
Decision-making skills	Baseline mean	142	4.90	1.15	139	4.94	1.20	–0.02	–0.02

Note: *N* represents the number reporting.

As shown in Table 8, the standardized differences in the baseline measure of the outcome for each analytic sample range from 0.01 to 0.07. All but one (perceived adult support) are within the WWC's range for satisfying baseline equivalence (between 0.00 and 0.05), and the remaining estimate is within the statistical adjustment range (between 0.05 and 0.25). Given that we include the baseline measure of the outcome as a covariate in our impact models, all of the analytic samples at this time point satisfy the WWC's baseline equivalence standards without the need for statistical weights.[31]

IMPACT ESTIMATES

Table 9 presents the benchmark impact estimates for each outcome assessed at the end of the first academic year (end of 10th grade). For each outcome, we present the number of participants in the analytic sample, the regression-adjusted mean outcomes for the AM and control groups, the impact

[30] We calculate a model-based difference by regressing the baseline measure of the outcome on the treatment indicator, as well as a series of dummy variables representing randomization blocks (school site and cohort). We calculate the standardized difference using the formula for Hedges' *g* as outlined in the WWC Standards and Procedures Handbook, Version 5.0 (WWC, 2022) when the baseline measure is continuous, and the Cox Index when the baseline measure is dichotomous.

[31] As shown in Table 7, the overall and attrition rates for all analytic samples at this time point are within the WWC's cautious boundary for potential threat of bias and we do not need to present baseline balance estimates for these samples, as a result. However, we do so out of an abundance of transparency.

coefficient estimate of interest and its standard error, as well as the *p*-value and effect size for the impact estimate. We describe results from each of the four research questions following Table 9.

Table 9. Impact Findings – 10th Grade[32]

Outcome Measure	Number Reporting	AM Mean	Control Mean	Mean Difference (SE)	*p*-value	Effect Size
Behavioral and academic						
Days attended in 10th grade	309	147.91	146.19	1.72 (2.19)	0.432	0.06
Credits earned in 10th grade	308	8.03	7.37	0.67 (0.51)	0.196	0.10
Suspended during 10th grade	302	20.06%	30.10%	–0.10 (0.05)	0.036*	–0.44
Social and emotional						
Perceived adult support	263	5.62	5.15	0.47 (0.16)	0.003**	0.37
Perceived peer acceptance	282	4.77	4.69	0.08 (0.13)	0.561	0.06
Academic self-concept	282	4.72	4.59	0.13 (0.09)	0.144	0.13
Academic self-efficacy	287	5.66	5.34	0.31 (0.13)	0.013*	0.26
Self-efficacy in help-seeking	287	5.49	5.16	0.33 (0.16)	0.037*	0.23
Self-efficacy in goal-setting	287	5.61	5.32	0.30 (0.14)	0.036*	0.22
Decision-making skills	281	5.24	5.02	0.22 (0.13)	0.087~	0.18

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

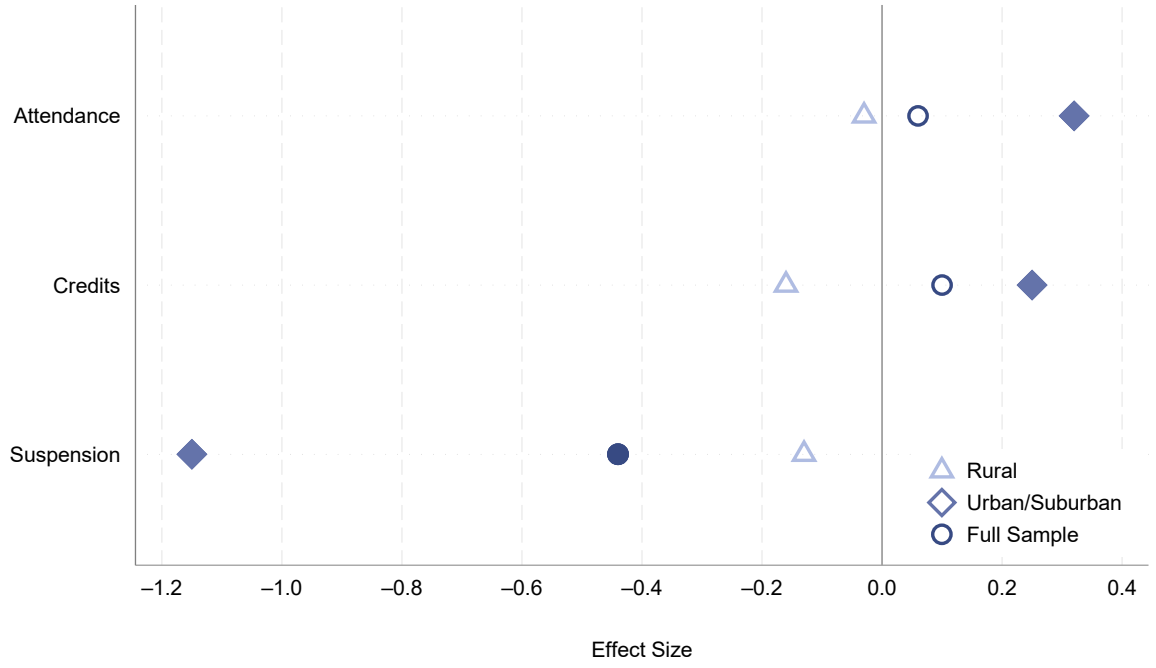
BEHAVIORAL AND ACADEMIC OUTCOMES

Benchmark statistical estimates indicate that offering one-on-one mentoring through the AM program for one year had a statistically significant effect on students' likelihood of getting suspended in 10th grade ($p = 0.036$), but not on their school attendance or credit accrual rate ($p > 0.10$). Specifically, model estimates presented in Table 9 show that 30.1% of control students were reportedly suspended during 10th grade compared with only 20.1% of students assigned to the AM condition, a statistically significant difference of 10% with effect size (Cox Index) of –0.44. Conversely, model estimates indicate that students assigned to the AM condition attended school an average of 1.7 days more than students assigned to the control condition (147.9 vs. 146.2 days, respectively), and earned two thirds of an additional credit than their control counterparts (8.0 vs. 7.4, respectively).

We conducted a series of subgroup analyses on each outcome to assess heterogeneity of observed benchmark effects. For each outcome, we examined treatment effect on subgroups of students based on the following baseline profiles: race/ethnicity, academic disadvantage, specific program eligibility criteria met, urbanicity, and students who self-report two or more barriers to school participation. One distinction that stood out as being different from main findings was urbanicity. Figure 4 presents the standardized treatment effect sizes for each outcome for the full analytic sample, and for the subgroup of students in either rural or urban/suburban schools.

[32] Model-adjusted group means and mean differences are calculated using an OLS model. Effect sizes for continuous outcomes are calculated using the formula for Hedges' *g*; we use the formula for the Cox Index for the dichotomous suspension outcome, which uses the difference in log odds generated using a logistic regression model.

Figure 4. *Academic and Behavioral Effect Sizes, Overall and by Urbanicity at End of 10th Grade*



Note: Symbols that are filled in denote estimates where $p < 0.10$.

As shown in Figure 4, we see that the positive trend in administrative outcomes observed for the full sample is driven primarily by students in urban/suburban schools. For example, we see that AM students in urban/suburban schools were significantly less likely than their control counterparts to get suspended in 10th grade ($p = 0.008$), whereas in rural schools, the finding was not significant. Moreover, among students who attended non-rural schools, we find that treatment students attended 5 more days of school ($p = 0.077$) and earned 2.5 more credits ($p = 0.077$) than their control counterparts. By contrast, in rural schools, AM students attended slightly fewer days of school and earned fewer credits than their control counterparts, though observed differences are not statistically significant ($p > 0.10$).[33] There were no other substantial findings of note across other subgroups.[34]

SOCIAL AND EMOTIONAL OUTCOMES

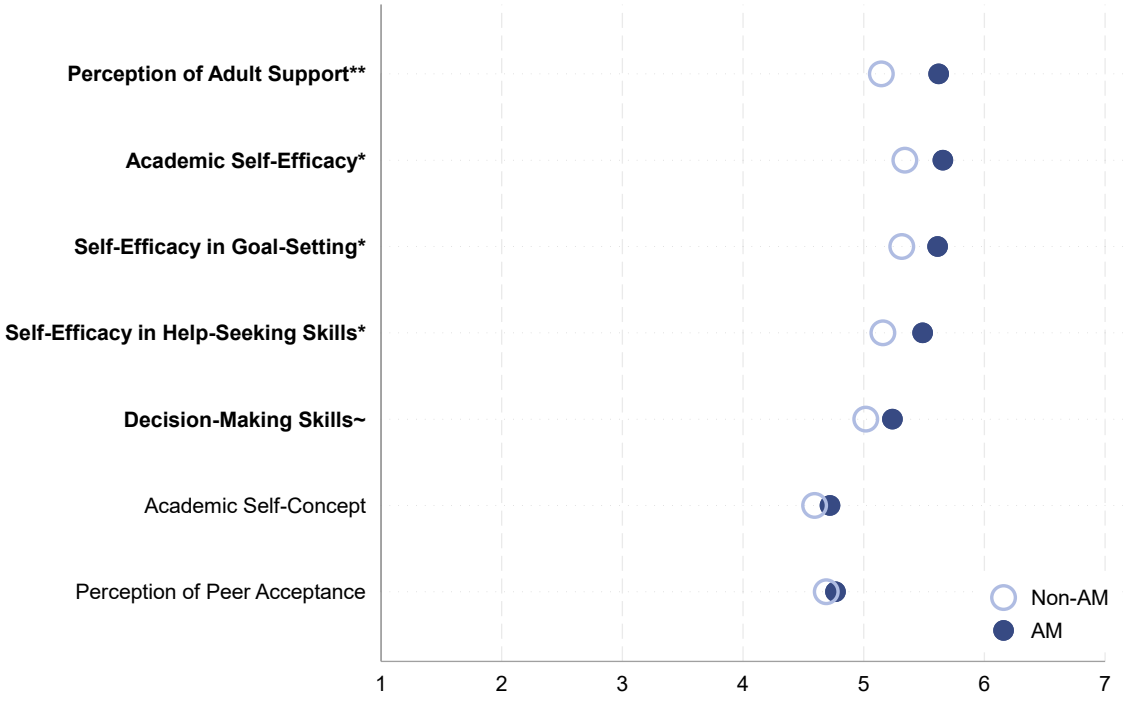
In terms of students' social and emotional outcomes, benchmark statistical estimates suggest that offering AM to 10th-grade students had a statistically significant effect on four of the seven outcomes and a marginally significant effect on one additional outcome measured at the end of one year. Specifically, model estimates (presented in Table 9) indicate students who were offered AM reported comparatively greater perceptions of adult support ($p = 0.003$), academic self-efficacy ($p = 0.013$), help-

[33] Figure 4 presents the effect size and statistical significance for impact estimates as modeled separately for each subgroup of students (either rural or urban). To examine the extent to which the differences between impact estimates for rural and urban schools were simply a factor of different sized samples, we also ran a model with the pooled sample and an interaction term (where treatment and urbanicity interacted). Findings from this pooled model corroborate our interpretation that the program is having a larger, more positive effect for students in urban/suburban schools compared with students in rural schools.

[34] The details of these subgroup and sensitivity analyses can be found in Appendix D.

seeking skills (p = 0.037), goal-setting behaviors (p = 0.036), and decision-making skills (p = 0.087) than students who were offered the control condition. The standardized magnitude of effect (Hedges' g) ranges from 0.18 to 0.37 on these measures. Model estimates do not support the hypothesis that offering AM has a statistically significant impact on students' academic self-concept or perceptions of acceptance by peers (p > 0.10). Figure 5 presents the regression-adjusted mean scale scores for the treatment (AM) and control groups on each measure; statistically significant differences are denoted with bolded labels.

Figure 5. Mean Scale Scores Reported at End of 10th Grade



Note: ~ p < 0.10; *p < 0.05; **p < 0.01.

Similar to the academic and behavioral outcomes, we examined treatment effects for subgroups of our sample based on baseline characteristics. We did not observe any meaningful deviations from the benchmark findings on social and emotional outcomes; on average students assigned to AM reported comparatively higher scores on measures than control students, with similar magnitudes of effect to what is reported in Table 9. Additional details can be found in Appendix D.

DISCUSSION OF 10TH GRADE FINDINGS

Benchmark findings at the end of the first time point (10th grade) provide some promising support for the hypothesis that offering regular, individualized mentoring to students who meet one or more risk indicators at the start of 10th grade can improve their academic, behavioral, and social and emotional outcomes. Model estimates indicate that students who were offered AM at the start of 10th grade self-reported feeling more supported by adults at school and more confident in their abilities to practice goal-setting, help-seeking, and academic skills. They also self-reported using critical decision-making skills more frequently, and administrative data indicate they were less likely to be suspended during

their 10th grade year. Although findings related to attendance, credit accrual, and academic self-image were not statistically significant, mean outcomes for AM students were consistently greater than their control counterparts.

In terms of students' social and emotional outcomes, AM students self-reported greater perceptions of support, a construct within the broader domain of school climate and safety, and self-efficacy. Standardized magnitude of effect (*g*) for the measure of perceived adult support was 0.37; effect size ranged from 0.22 to 0.26 for measures of self-efficacy. A recent systematic review of studies on universal SEL programs suggests reviewed programs had an average effect of 0.21 on attitudes and beliefs (95% confidence interval [CI] 0.16 to 0.26), including self-efficacy, and 0.29 on school climate and safety (CI 0.20 to 0.39) (Cipriano et al., 2023). The present findings suggest that offering one year of individualized mentoring to students who are at risk of dropping out after their first year of high school has comparable effects to the upper range of effects for these universal programs. In the same review, authors did not find a statistically significant effect of universal programming on disciplinary outcomes, whereas we observe a significant and moderately large effect on suspension rate when students are offered individual mentoring during their second year of high school (–0.44).

Subgroup analyses provide some evidence that the program may be more effective at improving administrative outcomes among students in non-rural schools. Specifically, students assigned to AM at urban or suburban schools earned more credits, attended more days of school, and were less likely to get suspended during 10th grade.

ELEVENTH GRADE

ATTRITION AND BASELINE EQUIVALENCE

Table 10 presents, for each outcome measure, the number of participants randomized into each treatment condition and that were eligible to contribute data at the second time point (i.e., were enrolled during the 2021–22 or 2022–23 school year), the number in each group that had data at the end of 11th grade, as well as the overall and differential attrition rates for the outcome and whether the combination was below the WWC's cautious boundary for an acceptable threat of bias due to attrition.

Table 10. Randomized and Analytic Samples – 11th Grade

	Number Randomized		Analytic Sample				
Outcome Measure	**AM**	**Control**	**AM**	**Control**	**Overall Attrition**	**Differential Attrition**	**Under WWC Cautious Boundary**
Days attended	165	164	95	93	42.9%	0.9%	Yes
Credits earned	165	164	93	92	43.8%	0.3%	Yes
Suspension rate	165	164	87	89	46.5%	–1.5%	Yes
Perceived adult support	165	164	87	91	45.9%	–2.8%	No
Perceived peer acceptance	165	164	88	94	44.7%	–4.0%	No
Academic self-concept	165	164	89	91	45.3%	–1.5%	Yes
Academic self-efficacy	165	164	90	95	43.8%	–3.4%	No
Self-efficacy in help-seeking	165	164	89	94	44.4%	–3.4%	No
Self-efficacy in goal-setting	165	164	87	93	45.3%	–4.0%	No
Decision-making skills	165	164	89	94	44.4%	–3.4%	No

As depicted in Table 10, at the second time point (end of 11th grade), overall attrition rates range between 43 and 47%, with differential rates ranging between 0 to 4%. A substantial portion of the overall attrition resulted from three study schools (one in Cohort 1 and two in Cohort 2) discontinuing participation in the project after their first year of implementation. As a result, three of our randomization blocks totaling 80 students are not represented in the final analytic samples. A second reason for higher attrition is due to our inability to gather data on/from students who left study schools or districts before the end of the second year.

The combination of overall and differential attrition rates for our three administrative outcomes (attendance, credits, and suspensions) as well as one self-report outcome (academic self-concept) are below the WWC's cautious boundary. The remaining six self-report outcomes, however, exceed this threshold and therefore have "unacceptable levels of potential bias." As a result, guidance indicates that baseline equivalence must be examined for analytic samples with high attrition and an acceptable statistical adjustment strategy (e.g., including the baseline measure of the outcome in the analytic model) should be used where differences exceed 0.05 standard deviation, but are below 0.25. One analytic sample, self-efficacy for goal-setting, has a baseline difference that exceeds this adjustment range and therefore we conducted a quasi-experimental propensity score weighting procedure to improve the baseline balance of the sample.[35] Table 11 presents the baseline balance estimates for each analytic sample at the 11th-grade time point, including the propensity-score weighted sample for self-efficacy in goal-setting.

[35] The unweighted standardized difference of means for the self-efficacy in goal-setting measure was 0.27 standard deviation. This exceeds the WWC's allowable threshold for baseline differences (0.25 standard deviation). As a result of failing to achieve low attrition and an acceptable baseline difference, we conducted a quasi-experimental propensity score weighting procedure to improve the balance of this analytic sample. Specifically, we first used a logistic regression model to generate an empirical score that quantifies the conditional probability that an individual would select into the treatment group – or alternatively the control group (i.e., the propensity score). Propensity scores were generated using baseline demographic data and baseline measures of all outcomes. We then use this propensity score to up-weight those cases that are more alike (i.e., are in the middle of the likelihood distribution) and down-weight those that are less alike (i.e., are on the extremes of the likelihood distribution). We then include the weight in the regression model that estimates baseline differences between the treatment and control participants in this analytic sample. We use the same weight in the impact analytic model for this outcome.

Table 11. Baseline Equivalence of Treatment and Control Groups – 11th Grade

		Treatment Group			Control Group			Treatment – Control	Standardized
Analytic Sample	**Baseline Measure**	***N***	**Mean**	**SD**	***N***	**Mean**	**SD**	**Difference**	**Difference**
Attendance	Days attended in 9th grade	95	145.32	26.17	93	146.29	24.92	–0.35	–0.01
Credits	Credits earned in 9th grade	93	9.51	8.98	92	8.67	8.32	0.07	0.01
Suspension	Percent suspended in 9th grade	87	17.24%	38.00	89	19.10%	39.53	–0.03	–0.12
Perceived adult support	Baseline mean	87	5.12	1.13	91	4.97	1.15	0.21	0.19
Perceived peer acceptance	Baseline mean	88	4.33	1.55	94	4.39	1.19	0.11	0.08
Academic self-concept	Baseline mean	89	4.61	1.02	91	4.47	1.03	0.15	0.14
Academic self-efficacy	Baseline mean	90	5.60	0.85	95	5.51	1.06	0.12	0.13
Self-efficacy in help-seeking	Baseline mean	89	5.23	1.27	94	5.04	1.22	0.29	0.23
Self-efficacy in goal-setting	Baseline mean	87	5.38	1.27	93	5.31	1.31	0.13	0.10
Decision-making skills	Baseline mean	89	4.82	1.20	94	4.86	1.20	0.08	0.07

As shown in Table 11, the standardized differences in the baseline measure of the outcome for each analytic sample range from 0.01 to 0.23. Given that we include the baseline measure of the outcome as a covariate in our impact models, all of the analytic samples at this time point satisfy the WWC's baseline equivalence criteria.

IMPACT ESTIMATES

Table 12 presents the impact estimates for each outcome assessed at the end of the second academic year (end of 11th grade). For each outcome, we present the number of participants in the analytic sample, the regression-adjusted mean outcomes for the AM and control groups, the impact estimate of interest and its standard error, as well as the *p*-value and effect size for the impact estimate.

Table 12. Impact Findings – End of 11th Grade[36]

Outcome Measure	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Behavioral and academic						
Days attended in 10th and 11th grades	188	294.02	303.46	–9.44 (6.77)	0.165	–0.17
Credits earned in 10th and 11th grades	185	25.05	24.22	0.83 (1.06)	0.432	0.06
Suspended during 10th or 11th grades	176	35.96%	33.75%	0.02 (0.07)	0.757	0.05
Social and emotional						
Perceived adult support	178	5.96	5.48	0.48 (0.23)	0.042*	0.35
Perceived peer acceptance	182	5.24	5.03	0.21 (0.21)	0.308	0.15
Academic self-concept	180	4.77	4.48	0.29 (0.12)	0.019*	0.29
Academic self-efficacy	185	5.66	5.56	0.10 (0.16)	0.535	0.09
Self-efficacy in help-seeking	183	5.35	5.27	0.08 (0.21)	0.704	0.06
Self-efficacy in goal-setting[37]	180	5.58	5.66	–0.08 (0.18)	0.654	–0.07
Decision-making skills	183	5.35	4.92	0.43 (0.19)	0.022*	0.35

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

BEHAVIORAL AND ACADEMIC OUTCOMES

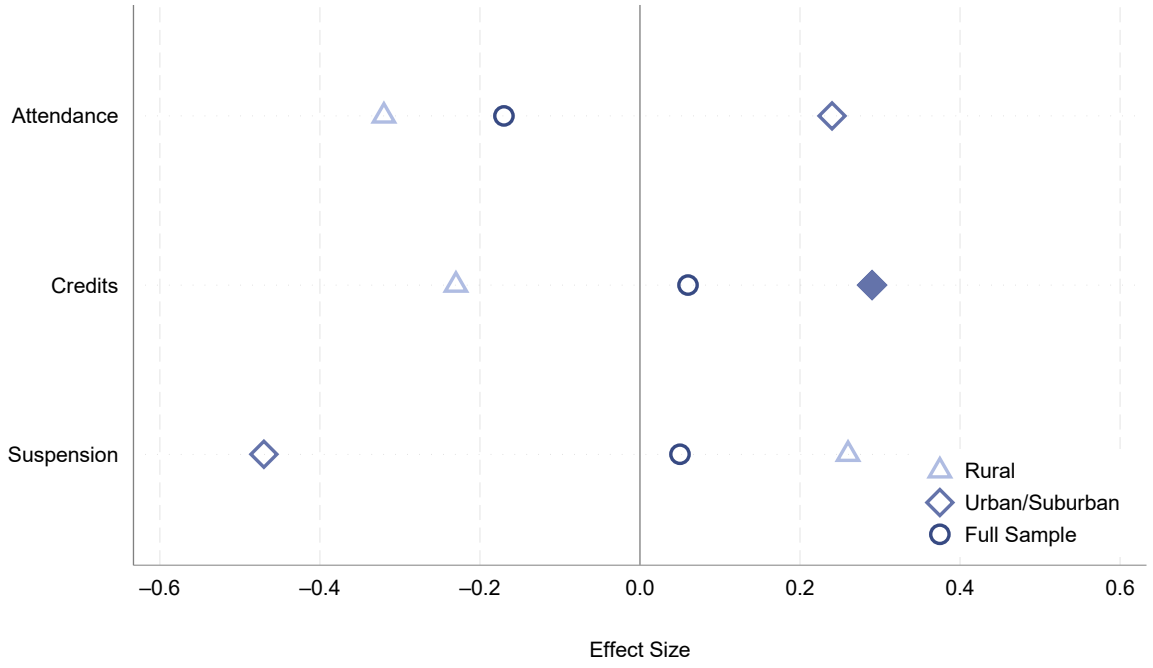
Benchmark statistical estimates indicate that offering two years of one-on-one mentoring through the AM program had no statistically significant effect on students' academic and behavioral outcomes, including number of days attending school, number of credits accrued, and likelihood of getting suspended, at the end of 11th grade. Model estimates presented in Table 12 show that students randomly assigned to AM earned a similar number of credits across 10th and 11th grades compared with their control counterparts (25 vs. 24 credits, respectively) and were similarly likely to get suspended in either grade (36% vs. 34%, respectively). Effect sizes for these outcomes were small (0.05 and 0.06). Model estimates indicate that students randomly assigned to the AM condition attended an average of 9 fewer days of school during 10th and 11th grades than their control counterparts, with a moderate, yet statistically insignificant ($p > 0.05$) effect size of –0.17.

We conducted a series of subgroup and sensitivity analyses on each outcome to assess heterogeneity of these null effects. For each outcome, we examined treatment effect on subgroups of students based on race/ethnicity, urbanicity, academic disadvantage, specific program eligibility criteria, and students who self-report two or more barriers to school at baseline. Again, we see a singular distinction warranting discussion. Figure 6 again presents a comparison of the standardized effect sizes for urban and rural subgroups, compared with the full sample.

[36] See footnote 32.

[37] When we exclude the propensity-score-generated sample weights from the analytic model, the impact coefficient and standard error are 0.00 and 0.18, respectively ($p = 0.980$).

Figure 6. *Academic and Behavioral Effect Sizes, Overall and by Urbanicity at End of 11th Grade*



Note: Symbols that are filled in denote estimates where $p < 0.10$.

As shown in Figure 6, we see that AM students in non-rural schools ($n = 44$) earned 7 additional credits in 10th and 11th grades ($p = 0.079$) compared with control students at these schools. Although not statistically significant, the trends presented in Figure 6 mirror those we observe at the 10th-grade time point (and presented in Figure 4) suggesting that the program may be more effective for students who attend non-rural schools. Specifically, we observe that AM students at non-rural schools attended 10 more days than their control counterparts and were 5% less likely to get suspended during 10th and 11th grades.[38] There were no other substantial findings of note across other subgroups.[39]

SOCIAL AND EMOTIONAL OUTCOMES

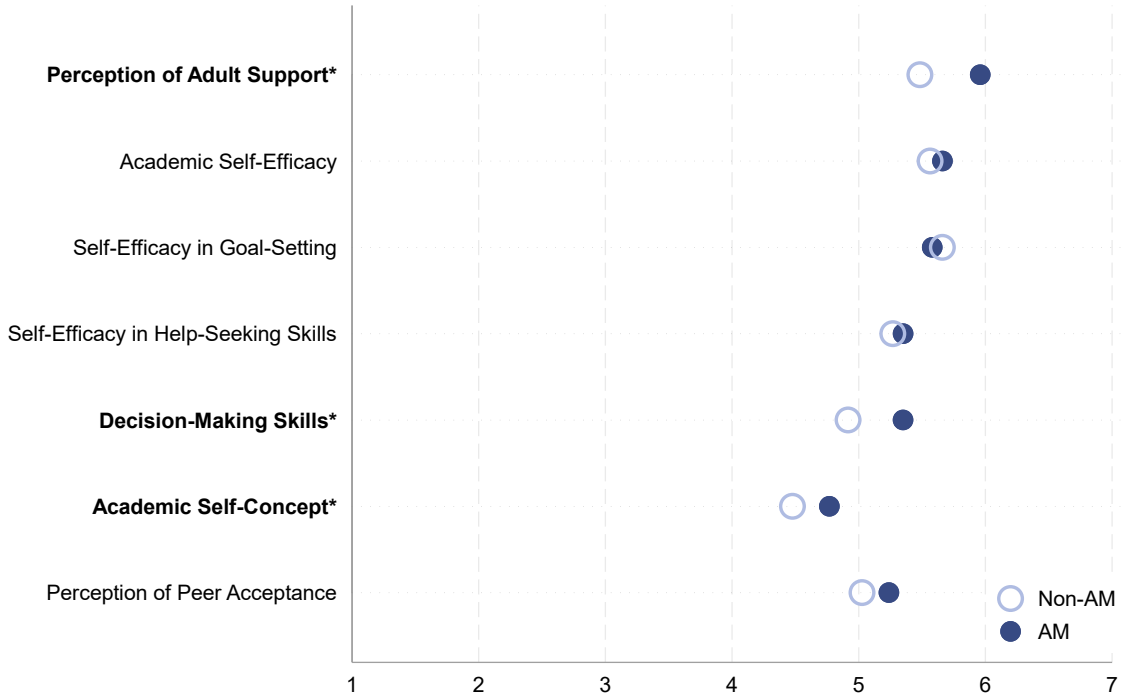
In terms of students' social and emotional outcomes, benchmark statistical estimates suggest that offering two years of AM to 10th-grade students had a statistically significant effect on three of the seven outcomes measured at the end of 11th grade. Figure 7 presents the regression-adjusted mean scale scores for the treatment (AM) and control groups on each measure; statistically significant differences are denoted with bolded labels. Specifically, model estimates indicate students who were offered AM reported comparatively higher mean scale scores on measures of perceived adult support ($p = 0.042$), academic self-concept ($p = 0.019$), and decision-making skills ($p = 0.022$) than students who were offered the control condition. The standardized magnitude of effect (Hedges' g) ranges from 0.29 to 0.35 on these measures. Model estimates do not support the hypothesis that offering two years of

[38] See footnote 33. Findings from a pooled model again corroborate our interpretation that the program is having a larger, more positive effect for students in urban/suburban schools compared with students in rural schools on the number of credits earned.

[39] See footnote 34.

AM has a statistically significant impact on students' self-efficacy or perceptions of acceptance by peers by the end of 11th grade ($p > 0.05$).

Figure 7. Mean Scale Scores Reported at End of 11th Grade[40]



Similar to the academic and behavioral outcomes, we examined treatment effects for subgroups of our sample based on baseline characteristics. We did not observe any meaningful deviations from the benchmark findings on social and emotional outcomes; on average students assigned to AM reported comparatively higher scores on measures of perceived support, self-concept, and decision-making skills than control students. Additional details can be found in Appendix D.

DISCUSSION OF 11TH-GRADE FINDINGS

Benchmark findings at the end of the second time point (11th grade) provide modest evidence that offering two years of individual mentoring to students identified as at risk for dropping out of school improves academic, behavioral, and social and emotional outcomes. In the full ITT sample for whom we have data, we do not observe significant impacts on administrative outcomes (attendance, credit accrual, discipline); however, we do see impacts on select social and emotional outcomes at the end of 11th grade. Model estimates indicate that students assigned to AM self-reported feeling more supported by adults at school, identifying as someone who could succeed academically, and practicing critical decision-making skills more frequently at the end of 11th grade ($p < 0.05$), but did not report greater self-efficacy or peer acceptance.

[40] ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

In terms of social and emotional outcomes, standardized magnitudes of effect (*g*) on significant outcomes ranged from 0.29 (academic self-concept) to 0.35 (perceived adult support and decision-making skills). Similar to 10th-grade findings, these effect sizes align with the meta-analytic findings described in Cipriano et al. (2023) for attitudes and beliefs (0.21) and school climate and safety domains (0.29), suggesting that offering two years of individualized mentoring to students identified as being at risk for dropping out of school may improve attitudes toward school above and beyond what a school-wide intervention might do.

Subgroup analyses provide some evidence that the program improved students' academic and behavioral outcomes in non-rural schools. Specifically, students assigned to AM at urban or suburban schools earned more credits, attended more days of school, and were less likely to get suspended during 10th and 11th grades. This again suggests that the program may be more effective at improving outcomes in non-rural settings, though given the very small analytic sample size (n = 44) at this time point, subsequent research is necessary to confirm this trend.

DISCUSSION

Results from our evaluation of AM provide promising evidence of the program's potential to improve outcomes for students identified as being at risk of dropping out of high school. We assessed the program's impact on student academic, behavioral, and social and emotional outcomes at two time points, once at the end of 10th grade (after one year of study enrollment) and again at the end of 11th grade (after two years of study enrollment) for the subsample of our ITT population who contributed outcome data. Empirical results at the end of 10th grade suggest that offering one year of AM to treatment students led to comparatively higher scores on measures of perceived support, self-efficacy, and decision-making skills, as well as a lower proportion of students getting suspended compared with control students who were not offered mentoring. Absolute values of standardized treatment effects (effect sizes) range from 0.18 to 0.44 across these outcomes, indicating a moderate to large effect by suggested educational benchmarks (Kraft, 2019). Although results were not significant on remaining outcomes of attendance, credits, and academic self-concept, students assigned to be offered AM reported comparatively favorable outcomes compared with control students, with effect sizes ranging from 0.06 to 0.13, indicating a small, yet potentially meaningful effect.

Although we expected program effects to persist after two years, findings are considerably more modest when we examine the effect of AM on 11th-grade outcomes for the subsample of our ITT population who contributed data. Considering students in the 11th-grade analytic sample, we find that those who were offered two years of AM reported comparatively higher scores on measures of perceived support, academic self-concept, and decision-making skills compared with control students who were not offered mentoring. Effect sizes for these measures ranged from 0.29 to 0.35, suggesting moderate to large effects on these outcomes (Kraft, 2019). In addition, similar to the positive trends we see at the end of 10th grade, AM students generally earned more credits than control students, though this difference is not statistically significant. The other promising program effects observed at the end of 10th grade (students assigned to AM had a lower suspension rate, attended more days of school, and self-reported higher levels of self-efficacy) are not seen at the end of two years. And though findings are not statistically significant, it is worth mentioning that in the case of attendance and suspensions, the overall positive trend in outcomes is not observed here – AM students are slightly more likely to get suspended and attended fewer days of school. We approach these potentially conflicting and puzzling findings at the end of 11th grade with caution given that we experienced a considerable level of overall and differential sample attrition and have a smaller sample size, which can be more susceptible to

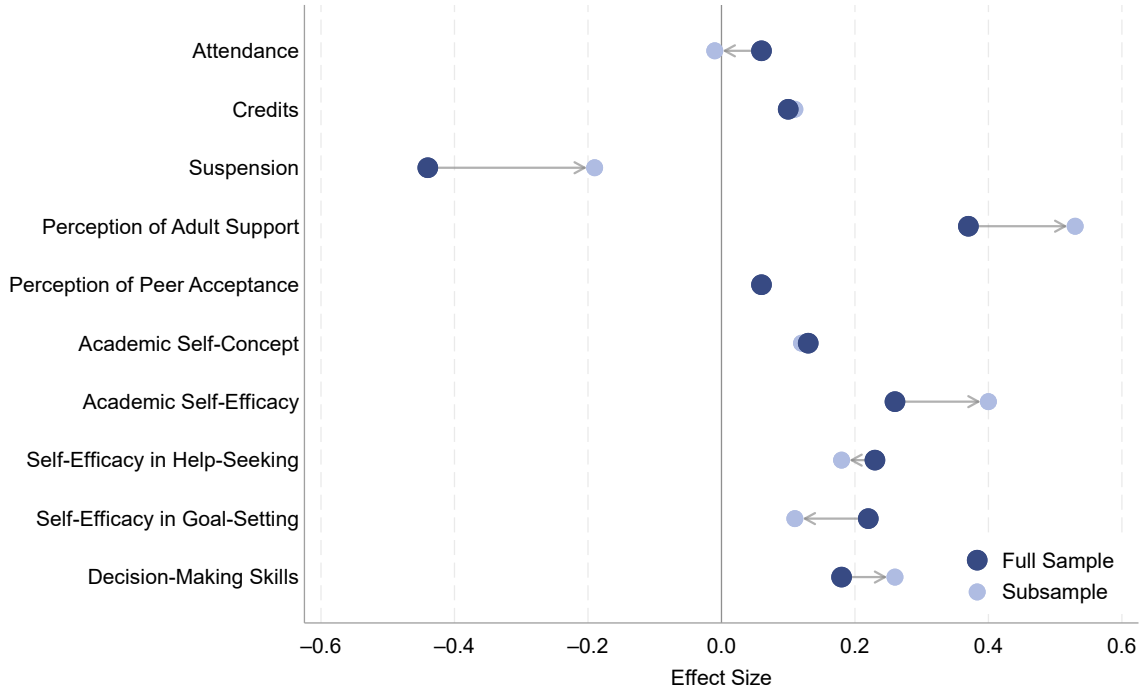
variability in measured outcomes than larger samples. Nevertheless, they warrant further exploration, in the event that they are evidence of real program effect trends.

We examined a few potentially explanatory hypotheses that could have explained why results are less robust at the second time point. The first question we explored was the extent to which simply having a smaller sample size, and therefore less statistical power to detect effects, accounts for the comparatively fewer significant results at the second time point. Standardized mean differences (effect sizes) reported in Tables 9 and 12 suggest, however, that smaller sample sizes do not necessarily account for the difference in findings at each time point, as we see smaller effect sizes on measures of self-efficacy (ranging from 0.06 to 0.09) at the latter time point compared with earlier (where these range from 0.22 to 0.26), and a smaller effect on suspensions (0.05 at 11th grade compared with –0.44 at 10th grade).

The second question we explored was the extent to which the students represented in our 11th-grade subsample actually received mentoring during their second year of participation, given that a substantial portion of our randomized sample attended schools that dropped out of the project. Implementation data do not support this explanation, however. Of the treatment students who are represented in at least one analytic sample (n = 114), 70% received at least one mentoring session during their second year in the study and on average, treatment students attended 14 sessions during 11th grade. Although this is below the recommended threshold of 20 sessions to meet fidelity of implementation, it is similar to the dosage level we observed during 10th grade.

The third question we examined was whether observed trends in outcomes may be sample specific. That is, if we examine 10th-grade outcomes only for the group of students who contributed 11th-grade data, do average 10th-grade findings remain, or are they potentially more similar to those observed for 11th grade? Data provide some support for the hypothesis that 10th-grade outcomes may be sample specific. Figure 8 presents the standardized effect size for each 10th-grade outcome for the full analytic sample for that time point (dark blue dots), and the subsample of students who contributed data at the second (11th-grade time point, lighter blue dots). Arrows represent the relative difference between these estimates as a means of presenting the degree to which findings for the two samples are similar or vary.

Figure 8. *Tenth-Grade Outcome Effect Sizes for Full and Subsamples*[41]



As shown in Figure 8, there is not a clear pattern in sample differences; in some cases, the standardized effect sizes for 10th-grade outcomes for the *subsample* are smaller (closer to 0) than those of the *full sample*, but in other cases, they are the same or larger. Specifically in terms of academic and behavioral outcomes, among students who contributed data at 11th grade, we see standardized effects of –0.01 for attendance, 0.11 for credits, and –0.19 for suspensions at 10th grade. Compared with the full sample, findings are consistent for credits (0.10), but considerably smaller for suspensions (–0.44) and attendance (0.06). For social and emotional outcomes, we see greater standardized effects on perceived support (0.53 for the subsample vs. 0.37 for the full sample), academic self-efficacy (0.40 vs. 0.26), and decision-making skills (0.26 vs. 0.18), but smaller effects for self-efficacy in help-seeking (0.18 vs. 0.23) and goal-setting (0.11 vs. 0.22). This suggests to us that the difference in composition between the full analytic sample assessed at 10th grade and the subsample of students who contributed data at 11th grade could be, at least partially, responsible for the inconsistencies we see in the program's effect across time.

Finally, we explored the extent to which there were variations in the program's observed effect across different subsamples of students as a means of better understanding where or for whom the program may be more or less effective. As discussed in the Impact Study Results section, we noted a potentially meaningful trend in the program's effect for students in urban and suburban schools compared with those in rural schools at both time points (Figures 4 and 6). Findings from these analyses suggest that the program may be more effective at promoting academic behaviors (attendance, suspensions, credits) for students in urban and suburban areas and less so for students in rural schools. We do still observe

[41] Model estimates for the subsample are provided in Table E.4 in Appendix E.

positive, significant effects for social and emotional outcomes for students who attended rural schools, suggesting the program does consistently improve attitudes and beliefs about their capacity to succeed. Although these trends are notable, we acknowledge that the sample sizes for these subgroup analyses are small and therefore warrant additional study to confirm.

IMPLICATIONS

Our evaluation has presented a broad range of data examining AM's impact on high school students' academic, behavioral, and social and emotional outcomes as hypothesized in the program's theory of change. The program's goal is to help students who are identified as being at risk for dropping out of high school ultimately make it to graduation. Although it was outside the scope of this evaluation to examine the program's impact on graduation rate directly, we contextualize our findings within the broader prevention literature below.

Empirical research strongly supports the relationship between getting suspended in high school and reduced achievement, as well as increased likelihood of dropout (Noltemeyer et al., 2015). Within our study, students who were identified as being at risk for dropping out and who were offered mentoring during 10th grade were 10% less likely to get suspended than similar students who were not offered mentoring.

Treatment students reported feeling more supported and practicing critical decision-making skills at the end of 10th and 11th grades. They also reported feeling more confident (self-efficacious) during 10th grade and saw themselves as the type of student who could succeed academically by the end of 11th grade. Self-efficacy, and in particular academic self-efficacy, has been shown to have a positive relationship with academic achievement and an inverse relationship with dropout (Bandura et al., 2001; Caprara et al., 2008; Caprara et al., 2011; Peguero & Shaffer, 2015;). Additional research indicates that decision-making skills are strongly predictive of increased academic performance (Dymnicki et al., 2013). Within our evaluation sample, however, we observe a weak relationship ($r < 0.30$) between measures of self-efficacy and decision-making skills and credit accumulation totals. Nevertheless, a vast body of work, led by the Collaborative for Academic, Social, and Emotional Learning, promote the development of social and emotional outcomes to improve life outcomes, such as academic performance, healthy relationships, and mental well-being.

The number of credits earned toward high school graduation is ultimately the most direct indicator of a student's likelihood of graduating with a high school diploma. Students who fail to accrue enough credits do not graduate. Although not statistically significant for the full sample, students in our study assigned to AM earned, on average, just under one additional course credit toward graduation at the end of 10th and 11th grades compared to their control counterparts.

We would argue that the results of this evaluation provide promising evidence that offering one-on-one adult mentoring through the AM model could improve the likelihood that students who are identified as being at risk of dropping out before graduation would graduate from high school. Although the findings from our multiyear evaluation were not always linear, taken together, these trends provide early support for the continued investigation of how AM could improve graduation rates for students who meet one or more risk indicators during ninth grade. It's clear that the AM program would benefit from subsequent, larger scale replication studies to continue to refine its logic model and investigate its effect on student outcomes more broadly and through the high school pathway to graduation.

LIMITATIONS

Limitations of this evaluation reflect common pitfalls in applied research. Although a well-executed RCT offers the most internally valid estimates of a program's impact, the design is not immune to external constraints. In this case, the primary constraint is sample loss due to attrition. The research team and grantee organization made every effort possible to retain as many study participants from our ITT sample into the analytic sample, but school- and district-level decisions to discontinue participation in the project increased our overall attrition rates beyond what we initially anticipated. In some cases (i.e., at the 11th-grade time point) the combination of overall and differential attrition exceeded the WWC's cautious boundary for acceptable threats of bias. Additionally, recruitment challenges in the immediate aftermath of the COVID-19 pandemic led to a small overall sample for the evaluation, producing weaker statistical power to detect programmatic effects than planned. We initially aimed to recruit 800 tenth-grade students into the ITT sample, enough to detect an effect size of 0.16 with statistical significance. Ultimately, with a final ITT sample of 393 students, we were only powered to detect a minimum effect of 0.22 (with fixed assumptions), before considering sample attrition. With an average analytic sample of 288 students at the 10th-grade time point and 180 at the 11th-grade time point, we were actually only powered to detect a minimal effect of 0.26 and 0.32, respectively. Future efforts to rigorously evaluate AM at a larger scale would improve our understanding of the program's theory of change and longer-term effects on student success and well-being.

REFERENCES

Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, *44*(5), 427–445. https://doi.org/10.1016/j.jsp.2006.04.002

Bandura, A. (1977). *Social learning theory*. Prentice-Hall.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. *Child Development*, *72*(1), 187–206. https://doi.org/10.1111/1467-8624.00273

Bathgate, M. E., Schunn, C. D., & Correnti, R. (2014). Children's motivation toward science across contexts, manner of interaction, and topic. *Science Education*, *98*(2), 189–215. https://doi.org/10.1002/sce.21095

Bowen, G. L., Richman, J. M., Bowen, N. K., & Broughton, A. (2003). The School Success Profile online. *Journal of Technology in Human Services*, *21*(1–2), 111–138. https://doi.org/10.1300/J017v21n01_06

Bruce, M., & Bridgeland, J. (January 2014). *The mentoring effect: Young people's perspectives on the outcomes and availability of mentoring*. Civic Enterprises with Hart Research Associates for MENTOR: The National Mentoring Partnership. https://www.mentoring.org/resource/the-mentoring-effect/

Bruce, M., Bridgeland, J. M., Fox, J. H., & Balfanz, R. (November 2011). *On track for success: The use of early warning indicator and intervention systems to build a grad nation*. Civic Enterprises.

Caprara, G. V., Fida, R., Vecchione, M., Del Bove, G., Vecchio, G. M., Barbaranelli, C., & Bandura, A. (2008). Longitudinal analysis of the role of perceived self-efficacy for self-regulated learning in academic continuance and achievement. *Journal of Educational Psychology*, *100*(3), 525–534. https://doi.org/10.1037/0022-0663.100.3.525

Caprara, G. V., Vecchione, M., Alessandri, G., Gerbino, M., & Barbaranelli, C. (2011). The contribution of personality traits and self-efficacy beliefs to academic achievement: A longitudinal study. *The British Journal of Educational Psychology*, *81*(Pt 1), 78–96. https://doi.org/10.1348/2044-8279.002004

Cipriano, C., Strambler, M. J., Naples, L. H., Ha, C., Kirk, M., Wood, M., Sehgal, K., Zieher, A. K., Eveleigh, A., McCarthy, M., Funaro, M., Ponnock, A., Chow, J. C., & Durlak, J. (2023). The state of evidence for social and emotional learning: A contemporary meta-analysis of universal school-based SEL interventions. *Child Development*, *94*(5), 1181–1204. https://doi.org/10.1111/cdev.13968

Dubois, D. L., Portillo, N., Rhodes, J. E., Silverthorn, N., & Valentine, J. C. (2011). How effective are mentoring programs for youth? A systematic assessment of the evidence. *Psychological Science in the Public Interest*, *12*(2), 57–91. https://doi.org/10.1177/1529100611414806

Dymnicki, A., Sambolt, M., & Kidron, Y. (March 2013). *Improving college and career readiness by incorporating social and emotional learning.* College and Career Readiness and Success Center. Retrieved from https://files.eric.ed.gov/fulltext/ED555695.pdf

Fall, A. M., & Roberts, G. (2012). High school dropouts: Interactions between social context, self-perceptions, school engagement, and student dropout. *Journal of Adolescence*, *35*(4), 787–798.

Goodenow C. (1993). The psychological sense of school membership among adolescents: Scale development and educational correlates. *Psychology in the Schools*, *30*(1)*,* 79–90. https://doi.org/10.1002/1520-6807(199301)30:1<79::AID-PITS2310300113>3.0.CO;2-X

Gordon, L. (2016, December 14). Mentoring's promise and limits. *The Atlantic*. https://www.theatlantic.com/education/archive/2016/12/why-some-mentors-fail/510467

Hanson, T. L., & Kim, J. O. (September 2007). *Measuring resilience and youth development: The psychometric properties of the Healthy Kids Survey* (Issues & Answers Report, REL 2007–No. 034). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory West. Retrieved from http://ies.ed.gov/ncee/edlabs

Johnson, W., McGue, M., & Iacono, W. G. (2006). Genetic and environmental influences on academic achievement trajectories during adolescence. *Developmental Psychology*, *42*(3), 514–532. https://doi.org/10.1037/0012-1649.42.3.514

Kraft, M. A. (2019). Interpreting Effect Sizes of Education Interventions (EdWorkingPaper: 19-10). Retrieved from Annenberg Institute at Brown University. http://www.edworkingpapers.com/ai19-10

McNeal, R. B., Jr., & Hansen, W. B. (1999). Developmental patterns associated with the onset of drug use: Changes in postulated mediators during adolescence. *Journal of Drug Issues*, *29*(2), 381-400. https://doi.org/10.1177/002204269902900214

National Center for Education Statistics. (2022). Table 233.40. Percentage of students suspended and expelled from public elementary and secondary schools, by sex, race/ethnicity, and state: 2017-18. *Digest of Education Statistics: 2022*. Retrieved from https://nces.ed.gov/programs/digest/d22/tables/dt22_233.40.asp

National Center for Education Statistics. (2024). High school graduation rates. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. Retrieved from https://nces.ed.gov/programs/coe/indicator/coi/high-school-graduation-rates

Noltemeyer, A. L., Ward, R. M., & Mcloughlin, C. (2015). Relationship between school suspension and student outcomes: A meta-analysis. *School Psychology Review*, *44*(2), 224–240.

Peguero, A. A., & Shaffer, K. A. (2015). Academic self-efficacy, dropping out, and the significance of inequality. *Sociological Spectrum*, *35*(1), 46–64. https://doi.org/10.1080/02732173.2014.978428

Pianta, R. C., Hamre, B. K., & Allen, J. P. (2012). Teacher-student relationships and engagement: Conceptualizing, measuring, and improving the capacity of classroom interactions. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 365–386). Springer.

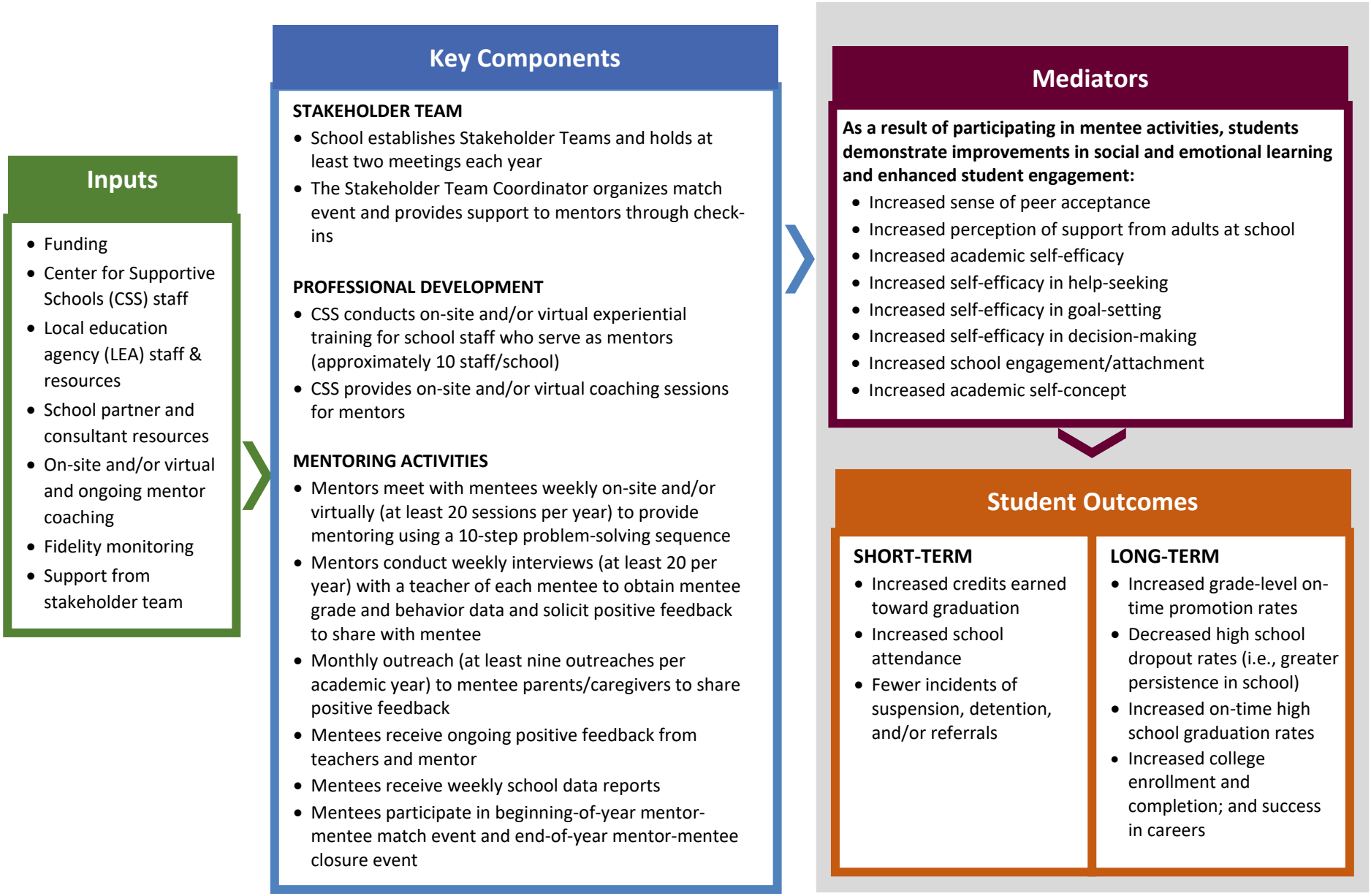
Stoker, G., Liu, F., & Arellano, B. (December 2017). *Understanding the role of noncognitive skills and school environments in students' transitions to high school* (REL 2018–282). U.S. Department of

Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs

What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook* (Version 5.0). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Handbooks

What Works Clearinghouse. (2024). *What Works Clearinghouse study review protocol* (Version 5.1). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Document/1297#/Protocol_CurrentProtocol

APPENDIX A. ACHIEVEMENT MENTORING LOGIC MODEL



APPENDIX B. FIDELITY OF IMPLEMENTATION

The purpose of this appendix is to present the results of PRG's fidelity evaluation of the Achievement Mentoring (AM) program, as required by the Education Innovation and Research (EIR) Program grant. The fidelity evaluation and reporting structure were developed by the EIR Evaluation Technical Assistance Team and provided to EIR evaluators to aid in cross-program examination of the extent to which funded interventions were implemented with fidelity. A detailed discussion of the formative implementation evaluation is provided in the Implementation Study Results section of this report.

AM was implemented in 12 schools over the period of three academic years (2021–22, 2022–23, and 2023–24). Schools are classified as *new* schools during their first year implementing AM and as *veteran* schools when they had implemented the program previously. We first present the research questions guiding the fidelity evaluation, followed by a description of the AM program, including its key components and the data sources used to monitor fidelity and the analysis approach. The fidelity matrices for each key component of AM are presented in Tables B.1 through B.3, followed by the aggregate sample-level implementation findings for each school year of implementation in Table B.4.

RESEARCH QUESTIONS

The fidelity study aimed to address one primary research question: Was each key component of the intervention implemented with fidelity? We also address two secondary research questions that examined the ways in which the model as implemented differed from the model as planned as well as identification of barriers and facilitators of implementation.

KEY COMPONENTS

STAKEHOLDER TEAM

The AM stakeholder team at each study school consists of four to six administrators, faculty, parents/guardians/caregivers, and students. The stakeholder team is led by a designated coordinator, who receives the training, tools, and resources necessary to meet regularly with the stakeholder team to plan and implement the program, troubleshoot obstacles, and ensure AM's long-term stability at the school. The Center for Supportive Schools (CSS) provides the stakeholder team with written protocols to select the mentors, which include resources for assessing qualifications and fit. During implementation, the stakeholder team provides support to mentors and the stakeholder team coordinator (STC), and advocates for the program in the school community. CSS typically facilitates two stakeholder team meetings, beginning in the planning period (during the spring/summer prior to AM implementation Year 1). One goal of these meetings is to build the capacity of the STC to facilitate regular team meetings throughout implementation and beyond. The STC acts as a liaison between the stakeholder team and CSS, coordinates communication and logistics for study data collection, and assists in obtaining school records and administrative data for ongoing monitoring of fidelity. The STC also organizes the mentor-mentee match process and kickoff event and provides support to mentors through brief check-in meetings to discuss program implementation and mentee progress.

PROFESSIONAL DEVELOPMENT

CSS provides initial and ongoing professional development to mentors throughout the two-year intervention. Selected mentors participate in a one-hour orientation session to learn about their role, followed by three days of (or the equivalent) intensive training over the course of the school year to learn how to serve in the role. CSS supports mentors through consultation/coaching meetings and on-

going technical assistance, provided on-site and/or virtually, to address challenging situations and to support progress, fidelity, and effectiveness of planning, preparation, and/or implementation as needed.

MENTORING ACTIVITIES

The stakeholder team selects mentors using resources for assessing qualifications and fit provided by CSS. The number of mentors and mentees at each school varies, depending on need and capacity. At the beginning of the first program year, mentor-mentee pairs participate in a beginning-of-the-year mentor-mentee matching event. During the school year, mentor-mentee pairs are expected to meet weekly for approximately 20-minute mentoring sessions (at least 20 sessions per year; conducted on-site or virtually). At the end of the school year, mentees and mentors participate in an end-of-the-year closure event.

Prior to each session, the mentor meets with one of the mentee's teachers to gather feedback on academic or classroom performance and behavior, along with positive feedback they can share with the mentee. The mentor uses this feedback to conduct the session using a structured, 10-step sequence:

1. Check in.
2. Review *Weekly Reporting Form* (WRF) and call attention to one or more instances of positive school engagement.
3. Praise specific instances of school engagement and encourage student to take credit for their instances of school engagement (i.e., ask, "How did you do it?").
4. Read aloud teacher's observations on WRF in objective, empathic manner.
5. Motivational interviewing, encourage students to talk about their need for change and their own reasons for wanting to change.
6. Refine mentee's SMART (specific, measurable, attainable, relevant, and time-bound) goal for the week.
7. Plan and practice implementation.
8. Students write the following week's goal on their WRF and return to class.
9. Mentor documents mentoring session using the *Weekly Online Mentoring Survey* (WOMS).
10. Mentor gives feedback to teachers, parents, and STC.

Mentors are expected to reach out to their mentee's parents/guardians once a month to share positive feedback about their student's academic behavior.

DATA COLLECTION PLAN AND KEY MEASURES

DATA SOURCES

WEEKLY ONLINE MENTORING SURVEY

The WOMS is completed by mentors after each planned mentoring session and collects information on program fidelity and dosage for each of their assigned mentees and mentoring sessions. The mentor responds *yes* or *no* to indicate whether they: (1) met with the mentee; (2) interviewed a teacher prior to the session; (3) shared the WRF with the mentee; (4) praised something about the mentee; (5) read mentee other teacher comments; (6) asked what the mentee made of the WRF; (7) helped the mentee refine a small goal for the week; and (8) planned a realistic implementation of the week's goal with the mentee. The mentor is expected to complete a WOMS each week regardless of whether they actually met with the mentee.

During the first two years of data collection (school years 2021–22 and 2022–23), WOMS were submitted through the IMPACT website. The IMPACT website was designed by a third party specifically

for the purpose of cross-site program monitoring, allowing for the extraction and presentation of data related to program implementation at the school and mentor level. Mentors and CSS coaches were able to log onto the website to enter or view recently entered data. Prior to the third year of implementation (2023–24), CSS discontinued use of the IMPACT website and created an online data entry form in Qualtrics. Mentors received personal links to this Qualtrics form via weekly email distribution set up by CSS that was linked to the *Implementation Tracking Tool*.

IMPLEMENTATION TRACKING TOOL

The *Implementation Tracking Tool* is a school-specific data collection tool developed by CSS. It is designed to house data on school-level aspects of program implementation. Data are entered on an ongoing basis throughout the planning and implementation period by the STC at each school and monitored for accuracy by CSS staff. Data recorded in the *Implementation Tracking Tool* include the name, contact information, and role of each member of the stakeholder team and each mentor, the dates of stakeholder team meetings, the dates of the kickoff and end-of-year match events, and the mentor-mentee pairs.

CSS PROGRAM MANAGER MEETING NOTES

PRG conducted biweekly meetings with CSS program managers from each region of study implementation (i.e., North Carolina, Delaware, Maryland, New Jersey, and Pennsylvania). During these meetings, a PRG analyst recorded information on whether any changes have been made to program content or implementation, including changes to mentor-mentee assignments, and any school- or community-wide events that occurred which may have an effect on implementation. During these calls, PRG also reviewed the WOMS response rate for each mentor and checked in with program managers about missing fidelity data, as needed.

ANALYSIS APPROACH

To assess the degree to which each key component of the intervention was implemented with fidelity, we reviewed data for each of the three intervention components, during each year of implementation. For each component, indicator scores were summed to create a total component score for each school. To determine whether a key component was implemented with fidelity for the full intervention sample, we calculate the percentage of schools that implemented the component with fidelity during each school year. We define the specific thresholds for implementation with fidelity – at both the school and sample levels – for each key component in Tables B.1 through B.3.

STAKEHOLDER TEAM

The Stakeholder Team key component fidelity is measured using five indicators. Schools with a score of 7 or higher are considered to have implemented the Stakeholder Team component with fidelity for the school year. The component was considered to have been implemented with fidelity at the sample level for the school year if 75% of schools implemented the component with fidelity. See Table B.1 for details.

Table B.1. *Key Component 1: Stakeholder Team*

Indicators	Definition	Unit of Implementation	Data Source	Score for Level of Implementation at Unit Level
1.1. Stakeholder team established	School recruits 4 to 6 school administrators and/or staff for stakeholder team	School	*Implementation Tracking Tool*	0 = no stakeholder team established 1 = stakeholder team established
1.2 Stakeholder team meetings	School holds at least two stakeholder team meetings each year	School	*Implementation Tracking Tool*	0 = no stakeholder team meetings held 1 = 1 stakeholder team meeting held 2 = 2 or more stakeholder team meetings held
1.3 Stakeholder team coordinator identified	School identifies a stakeholder team coordinator	School	CSS program manager meeting notes	0 = no stakeholder team coordinator identified 1 = stakeholder team coordinator identified
1.4. Stakeholder team coordinator support to mentors	Stakeholder team coordinator holds check-in meetings with mentors	School	*Implementation Tracking Tool* and/or WOMS	0 = no mentors report meeting with the stakeholder team coordinator 1 = at least one mentor reports checking in with their stakeholder team coordinator
1.5 Stakeholder team coordinator organizes match events	Stakeholder team coordinator organizes two mentee-mentor match events each academic year	School	*Implementation Tracking Tool*	0 = stakeholder team coordinator organized 0 events 1 = stakeholder team coordinator organized 1 event 2 = stakeholder team coordinator organized 2 events
All Indicators	Score range: 0–7 Unit-level adequate implementation score: 7			Adequate implementation at sample level: 75% of schools with score of 7

PROFESSIONAL DEVELOPMENT

The Professional Development key component fidelity is measured using two indicators. Indicator 2.1 is measured differently depending if the school is implementing the first year of AM or is implementing the second year of the program. New schools with a score of 5 are considered to have implemented the Professional Development component with fidelity for the school year; returning schools with a score of 2 are considered to have implemented the component with fidelity. The component was considered to have been implemented with fidelity at the sample level for the school year if 75% of schools implemented the component with fidelity. See Table B.2 for details.

Table B.2. *Key Component 2: Professional Development*

Indicators	Definition	Unit of Implementation	Data Source	Score for Level of Implementation at Unit Level
2.1a. Experiential training offered (Year 1 only)	CSS offers 4 experiential training sessions to mentors	School	Training attendance records	0 = CSS offered 0 training sessions 1 = CSS offered 1 training session 2 = CSS offered 2 training sessions 3 = CSS offered 3 training sessions 4 = CSS offered 4 training sessions
2.1b. Refresher training offered for veteran mentors (Year 2 only)	CSS offers refresher training to veteran mentors	School	Training attendance records	0 = CSS did not offer a refresher training 1 = CSS offered a refresher training
2.2. Virtual and/or on-site coaching	CSS offers coaching sessions, conducted in-person or virtually, to mentors	School	*Implementation Tracking Tool* and/or WOMS	0 = CSS provided 0 coaching sessions to mentors 1 = CSS provided at least 1 coaching session to mentors
All Indicators	Score range: 0–5 (Y1), 0–2 (Y2) Adequate implementation score: 5 (Y1), 2 (Y2)			Adequate implementation at sample level: 75% of schools with adequate implementation scores

MENTORING ACTIVITIES

The Mentoring Activities key component fidelity is measured using five indicators. Mentor-mentee pairs with a score of 10 or greater are considered to have implemented the Mentoring Activities component with fidelity. Mentor-mentee scores were rolled up to the school level and the component was considered to have been implemented with fidelity at the school level if at least 75% of mentor-mentee pairs achieved a score of 10 or greater. These school-level scores were again rolled up to the sample level and the component was considered to have been implemented with fidelity in the sample if 75% of intervention schools implemented the component with fidelity. See Table B.3 for details.

Table B.3. *Key Component 3: Mentoring Activities*

Indicators	Definition	Unit of Implementation	Data Source	Score for Level of Implementation at Unit Level
3.1. Mentors conduct weekly interviews with teachers	Mentors interview a teacher of the mentee before at least 70% of sessions	Mentor-mentee pair	WOMS	0 = mentors conduct interviews for 0–20% of mentoring sessions, on average 1 = mentors conduct interviews for 21–45% of sessions 2 = mentors conduct interviews for 46–69% of sessions 3 = mentors conduct interviews for 70–100% of sessions
3.2. Mentors meet with mentees	Mentors hold a minimum of 20 sessions with each mentee	Mentor-mentee pair	WOMS	0 = mentors meet 0–4 times with mentees, on average 1 = mentors meet 5–9 times with mentees 2 = mentors meet 10–14 times with mentees 3 = mentors meet 15–19 times with mentees 4 = mentors meet 20+ times with mentees
3.3. Mentor outreach to mentee parents/ guardians	Mentors conduct monthly outreach to parents/ guardians of each mentee	Mentor-mentee pair	WOMS	0 = mentors conduct 0–2 parent outreaches per mentee, on average 1 = mentors conduct 3–5 parent outreaches 2 = mentors conduct 6–8 parent outreaches 3 = mentors conduct 9+ parent outreaches
3.4. Mentees receive positive feedback from mentor	Mentees receive positive feedback from mentors at each session	Mentor-mentee pair	WOMS	0 = mentees receive positive feedback during less than 100% of sessions 1 = mentees receive positive feedback during 100% sessions
3.5. Mentees set a goal during mentoring sessions	Mentees set a goal at each mentoring session	Mentor-mentee pair	WOMS	0 = mentees set a goal at 0–35% of mentoring sessions, on average 1 = mentees set a goal at 36–69% of sessions 2 = mentees set a goal at 70–100% of sessions
All Indicators	Score range: 0–13 Adequate implementation score: 10			School level: Implemented with adequate fidelity = 75% or more of pairs were at adequate implementation Adequate implementation at sample level: 75% of schools with adequate implementation

RESULTS

Table B.4 presents the sample-level fidelity scores for each of the three key components during each study year, followed by a discussion of fidelity of program implementation.

Table B.4. AM Implementation Fidelity Findings

	Key Component 1: Stakeholder Team	Key Component 2: Professional Development	Key Component 3: Mentoring Activities
Year 1 (SY 2021–22)			
Percent of schools that met adequate implementation threshold	60% (3 of 5)	100% (5 of 5)	0% (0 of 5)
Sample met fidelity	No	Yes	No
Year 2 (SY 2022–23)[42]			
Percent of schools that met adequate implementation threshold	78% (7 of 9)	78% (7 of 9)	22% (2 of 9)
Sample met fidelity	Yes	Yes	No
Year 3 (SY 2023–24)[43]			
Percent of schools that met adequate implementation threshold	30% (3 of 10)	90% (9 of 10)	0% (0 of 10)
Sample met fidelity	No	Yes	No

Note: Samples met fidelity if 75% of implementing schools achieved an adequate implementation score for the given component.

As shown in Table B.4, the AM program was implemented with varying degrees of fidelity each school year of the study. Key Component 1 (Stakeholder Team) was implemented with fidelity during only one of the three implementation years (2022–23), whereas Professional Development (Key Component 2) was implemented with fidelity during all three school years. Key Component 3 (Mentoring Activities) was not implemented with fidelity during any school year during the evaluation.

Regarding the Stakeholder Team component, all but one school formed a stakeholder team and identified a STC to oversee program implementation (Indicators 1.1 and 1.3, respectively). Similarly, all but one school reported holding both the kickoff and closure events (Indicator 1.5), and the STC reported meeting with mentors during the school year (Indicator 1.4). Schools were less successful holding the recommended minimum number of stakeholder team meetings (Indicator 1.2) during the school year. Although it is possible that these meetings were held at the school and were simply not documented, PRG made every effort to triangulate information from the *Implementation Tracking Tool*, biweekly meeting notes, and email correspondence with the CSS coaches to confirm whether stakeholder teams met during each implementation year. The data available suggest that 58% of sites (14 of 24 implementation units across the three years) held two or more stakeholder team meetings.

[42] One school that participated during 2021–22 did not return to implement the program for a second year during 2022–23. We omit this school from the denominator when calculating the sample-level fidelity during 2022–23.

[43] One school that participated during 2022–23 did not return to implement the program for a second year during 2023–24. We omit this school from the denominator when calculating the sample-level fidelity during 2023–24.

CSS and schools were most successful in implementing professional development activities during each study year. In all three years, the majority (75% or greater) of schools were offered the recommended number of trainings (either 4 trainings in Year 1 or a refresher training in Year 2) and mentors reported meeting with their CSS coach at least once during the school year. Fidelity documentation suggests that during the 2022–23 school year, one school that was implementing the program for the first time was offered three of the four trainings whereas at another school, no mentors reported meeting with their CSS coach.

Weekly mentoring data suggest that schools experienced the most challenges implementing the weekly mentoring session activities with fidelity. As shown in Table B.4, only two schools during the 2022–23 school year implemented the component with fidelity. Indicator-level data assessed across the three years suggest that mentors were most consistent in facilitating weekly goal-setting with their mentees (67% of schools implemented Indicator 3.5 with fidelity). In addition, data indicate that approximately one third of schools implemented activities related to mentors interviewing their mentee's teachers (33%, Indicator 3.1) and mentees receiving positive feedback (29%, Indicator 3.4) with fidelity. Mentors were least successful in meeting with their mentees the recommended minimum of 20 times (Indicator 3.2) and contacting parents/guardians at least 6 times during the school year (Indicator 3.3). Specifically, mentors reported contacting parents between 4 and 5 times during a given school year and meeting with their mentees between 13 and 14 times.

APPENDIX C. IMPACT STUDY METHODS

The purpose of this appendix is to provide additional details of the impact study methods and data used to answer impact research questions. The impact study aimed to isolate the causal impact the Achievement Mentoring (AM) program had on 10th- and 11th-grade students' discipline, progression in school, attendance, and social and emotional outcomes. AM was designed to improve students' social and emotional learning skills and educational mindsets, thereby improving educational outcomes and behaviors. The target population for AM was incoming 10th-grade students identified as being at risk of dropping out of high school. The impact study was a randomized controlled trial. Outcomes for treatment group members who were offered the AM intervention were compared with those of a control group who were offered class as usual.

In this appendix, we provide additional details on the individual eligibility criteria for the study, outcome and covariate operationalization, analytic approach, and the methods used to establish baseline equivalence between the treatment and control groups.

DETAILED ELIGIBILITY CRITERIA

The study examines the effects of AM for 10th- and 11th-grade students in selected public high schools in rural North Carolina and urban communities in the mid-Atlantic region of the United States. The Center for Supportive Schools (CSS), the grantee and program developer, was responsible for selection, recruitment, confirmation, and retention of study high schools. CSS identified 13 schools in Maryland, New Jersey, Pennsylvania, Delaware, and rural North Carolina to participate in this study for three study cohorts. All participating schools serve large numbers of students representing subpopulations at disproportionate risk for poor academic outcomes, including exclusionary discipline practices (i.e., suspensions and expulsions). For each school that expressed interest in study participation, CSS conducted a readiness assessment prior to confirmation. CSS provided each study school with a fully developed curriculum, staff training, technical support, and financial support for the two years that the school implemented AM and participated in the study.

The Policy & Research Group (PRG) coordinated with school staff to screen all incoming 10th graders at the start of the academic year to determine whether or not each student on the school's 10th-grade roster met study eligibility criteria. PRG created an *Eligibility Screening Tool* to house each school's student-level eligibility data based on these criteria. Study schools were provided with two options for screening students for eligibility using this tool. To complete the screening process, study schools could either: (1) provide relevant administrative data to PRG, whereby PRG completed the *Eligibility Screening Tool*; or (2) fill out the *Eligibility Screening Tool* using administrative data and submit the completed spreadsheet to PRG. In all cases, schools submitted eligibility data through a secure file sharing system managed by PRG.

To be eligible for enrollment in the study, students had to meet all five of the following criteria:

1. Be enrolled in 10th grade at a study school at the time of randomization
2. Not already be enrolled in the study
3. Be considered at risk for dropping out of high school by meeting one or more of the following performance, attendance, and/or disciplinary criteria:
    a. Failed one or more core courses in the previous school year
    b. Missed more than 20 days in the previous school year, but attended school 3 days per week on average

   c. Incurred three or more disciplinary infractions (discipline referrals, suspensions, or detentions) in the previous school year
4. Provide consent (either passive or active) to participate in the study
5. Not be simultaneously enrolled in a similar youth development program (e.g., Peer Group Connection-High School)

We next describe each of the inclusion criteria in detail.

BE ENROLLED IN 10TH GRADE AT A STUDY SCHOOL

Students were required to be enrolled in 10th grade at a study school to participate in the study. If a student was on the roster of 10th graders provided by the school at the time of randomization, they were considered enrolled at that study school. To minimize attrition due to student mobility (e.g., transferring schools) and nonattendance, at the time of randomization, schools were asked to provide the most up-to-date roster of 10th-grade students possible before randomization, such that the up-to-date roster included only those students whose enrollment had been confirmed by their attendance in class or related school events. If it was determined that a study participant, for whatever reason, was no longer enrolled at the study school, but their departure date was after randomization, that student remained in the study and PRG attempted to collect questionnaire and student record data for the participant.

NOT ALREADY BE ENROLLED IN THE STUDY

Students could not be newly enrolled in the study during a given year if they were already enrolled in the study during a previous school year. This criterion applied when a school contributed more than one cohort of students to the study and a student was held back to repeat 10th grade. If a student on the roster of 10th graders provided by the school at the time of randomization had previously been randomized to either the treatment group or control group as part of an earlier study cohort, the student was deemed ineligible for the current year's cohort.

BE CONSIDERED AT RISK FOR DROPPING OUT OF HIGH SCHOOL

Students were required to be considered at risk for dropping out of high school in order to participate in AM. The program defines a student as at risk for dropping out if they meet one or more of the following criteria:

1. Failed one or more core classes (Math, Science, English, or Social Studies) in the previous school year
2. Missed more than 20 days in the previous school year, but attended at least 3 days per week on average
3. Incurred three or more of any combination of disciplinary infractions, such as discipline referrals, detentions, or suspensions, in the previous school year

Prior to randomization, school staff or PRG analysts used school records data from the previous school year (e.g., ninth grade) to determine if students met one or more of these criteria, and if so which one(s).

CONSENT TO STUDY PARTICIPATION

Students and parents were provided with an opportunity to consent to participation in the study prior to randomization at each school. Students and/or parents who did not consent to study participation were deemed ineligible for study participation. This eligibility criterion was assessed by schools prior to

randomization. The consent process varied among study schools; while some schools required active consent, others allowed for a passive (i.e., opt-out) process.[44]

ACTIVE CONSENT

Seven of the 13 study schools required active consent for students to participate in research activities.[45] For these schools, PRG provided consent forms in both English and Spanish to schools. School staff were then responsible for distributing consent forms (e.g., sending home forms through information packets, emailing parents a link to the form in Qualtrics) and following up with reminders to submit the form through meetings with students, emails and phone calls to parents, and other communication methods as needed.

PASSIVE CONSENT

Six of the 13 study schools did not require active consent for study participation. In these schools, consent was assumed for students unless a student or parent expressed a desire to opt out of the study. All students and parents at study schools were provided with study information and the opportunity to opt out prior to randomization. If a student opted out of the study following randomization, they were included in the study and considered a part of the intent-to-treat (ITT) sample but excluded from subsequent data collection efforts and the analytic sample (i.e., the case will contribute to study attrition).[46]

NOT BE SIMULTANEOUSLY ENROLLED IN A SIMILAR YOUTH DEVELOPMENT PROGRAM[47]

Students who were enrolled in any similar youth development or mentorship programming at the time of randomization were ineligible to participate in this study. School administrators were responsible for assessing this criterion; they reviewed the criterion using student records (e.g., student schedules). Students who were identified as participants of an alternative youth development program prior to randomization were excluded from eligibility for the study. Students who initiated participation in an alternative program after randomization remained in the study. PRG made every effort to communicate this eligibility criterion to the school partners to identify and document any widespread programming offered at the school.

VARIABLE OPERATIONALIZATION

In this section, we present a description of the individual-level covariates and the outcome variables used in the confirmatory impact analyses.

COVARIATES

Table C.1 provides a description of the individual-level covariates that were included in the Benchmark Analytic Model. The completeness of covariate data varied across individual variables. For the following

[44] PRG applied for an Institutional Review Board exemption on the grounds that AM is considered normal educational practice and research was conducted in established educational settings. PRG received the exemption and active consent was not required for the study. Some study schools and districts, however, required active consent for any research participation. In either case, PRG was responsible for managing the consent process and assessing the consent criterion prior to randomization.

[45] One study school switched from an opt-out system to active parental consent between Cohort 1 and Cohort 2.

[46] Over the course of the study, two students opted out of study participation following randomization.

[47] Many of the schools participating in this study had preexisting relationships with CSS and offered Peer Group Connection-High School (PGC-HS), an engagement intervention for all 9th-grade students scheduled into the school day. If the study enrolled any repeat 9th graders (students who were not promoted to Grade 10 at the end of their 9th-grade year), at the time of randomization, the school was required to confirm that they were not enrolled in PGC-HS and would not participate in PGC-HS activities. This criterion also applied to other structured youth development programs that might be offered to 10th and 11th graders at a given school.

covariates, data missing from school administrative records were imputed using the self-reported information provided by the student on the baseline *Participant Questionnaire*: age at baseline, race/ethnicity, gender. If covariate data were also missing on the baseline questionnaire, missing data were imputed using dummy variable adjustment.[48] All covariates were centered at the grand mean.

Table C.1. Covariate Operationalization

Variable Name	Variable Type, Construction, and Data Source
Age at baseline	Continuous – calculated by subtracting the student's date of birth from the date of the first day of school of the participant's 10th-grade year and dividing the difference by 365. **Data Source:** Baseline data request (school administrative records)
Race	A set of 4-1 mutually exclusive dummy (0/1) variables indicating the student's race/ethnicity. We include dummy variables representing students who were identified as: • White (1) or not (0) • Black or African American (1) or not (0) • Multiracial (1) or not (0) • Either Asian, Native American/Alaska Native, or Pacific Islander (1) or not (0) **Data Source:** Baseline data request (school administrative records)
Hispanic/Latino/a	A dummy (0/1) variable indicating the student identified as Hispanic/Latino/a (1) or not (0). **Data Source:** Baseline data request (school administrative records)
Gender[49]	Dummy variable indicating a student's gender as female (1) or male (0). **Data Source:** Baseline data request (school administrative records)
English Language Learner (ELL) status	Dummy variable indicating whether the student is designated as an ELL at the beginning of 10th grade (1) or not (0). **Data Source:** Baseline data request (school administrative records)
Individualized Education Plan (IEP) status	Dummy variable indicating whether the student has an IEP at the beginning of 10th grade (1) or not (0). **Data Source:** Baseline data request (school administrative records)
Number of days attended in 9th grade	Continuous – the number of days the student attended school during their 9th grade year (the year immediately preceding their participation in the study). Variable included in the Benchmark Analytic Model for Research Question 1 and represents the baseline measure of the outcome. **Data Source:** Baseline data request (school administrative records)
Number of credits earned in 9th grade	Continuous – the number of credits the student earned during their 9th grade school year (the year immediately preceding their participation in the study). Variable included in the Benchmark Analytic Model for Research Question 2 and represents the baseline measure of the outcome. **Data Source:** Baseline data request (school administrative records)
Number of days suspended in 9th grade	Continuous – the number of days the student was suspended during their 9th grade school year (the year immediately preceding their participation in the study). Variable included in the Benchmark Analytic Model for Research Question 3 and represents the baseline measure of the outcome. **Data Source:** Baseline data request (school administrative records)
Randomization blocks	A series of 18-1 dummy variables indicating whether the student was enrolled at a study school during a specified study cohort (1) or not (0). **Data Source:** Study Roster

[48] Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). *What to do when data are missing in group randomized controlled trials* (NCEE 2009-0049). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

[49] Study schools/districts reported gender along a binary categorization.

OUTCOME VARIABLES

Table C.2 outlines how measures for the three outcome research questions were constructed. Outcome measures assessed at the end of 11th grade were cumulative sums of data reported in 10th and 11th grades.

Table C.2. Outcome Variable Operationalization: Research Questions 1–3

Variable Name	Variable Type, Construction, and Data Source
School Attendance Number of days attended during 10th and 11th grade	Continuous: a. The number of days the student attended school during their regular 10th-grade year (end of 10th grade time point) b. The number of days the student attended school during their regular 10th- and 11th-grade years (end of 11th-grade time point) Only students with complete attendance data for the full school years are included in the analytic sample. Students who dropped out of school (W2 withdrawal) and did not transfer to another school were included in the analytic sample if the study school reported their attendance for the period they were enrolled in school (i.e., data were not missing). Students who transferred to a new school during 10th or 11th grade are included in the analytic sample if attendance data from all schools they attended were reported to PRG. Students who transferred to a new school and whose new school did not report their attendance were excluded from the benchmark analytic sample because their attendance was only partially reported for the year. Missing values were not imputed. **Data Source:** 10th- and 11th-grade data request (school administrative records)
Credits Earned Number of credits earned during 10th and 11th grade	Continuous: a. The number of credits earned at the end of their 10th-grade year (end of 10th grade time point) b. The number of credits earned at the end of their 10th- and 11th-grade school years (end of 11th-grade time point) If a student transferred to a new school during their 10th- or 11th-grade year, we requested that the new school report the total number of credits earned as of the end of the regular school year. Missing values were not imputed. **Data Source:** 10th- and 11th-grade data request (school administrative records)
Suspensions Number of days suspended during 10th and 11th grade	Dichotomous: a. A dummy variable indicating whether (1) or not (0) the student was suspended during their 10th-grade year (end of 10th-grade time point) b. A dummy variable indicating whether (1) or not (0) the student was suspended during their 10th- or 11th-grade years (end of 10th-grade time point) Only students with complete enrollment and discipline data for the full school years are included in the analytic sample. Students who dropped out of school (W2 withdrawal) and did not transfer to another school were excluded in the analytic sample. Students who transferred to a new school during 10th or 11th grade are included in the analytic sample if discipline data from all schools they attended were reported to PRG. Students who transferred to a new school and whose new school did not report their discipline data were excluded from the benchmark analytic sample because their discipline data was only partially reported for the year. Missing values were not imputed. **Data Source:** 10th- and 11th-grade data request (school administrative records)

QUESTIONNAIRE SCALE DETAILS

Table C.3 presents details of the outcome measures collected using the *Participant Questionnaire*, which was administered to study participants at the beginning and end of 10th grade and again at the end of 11th grade. We present, for each scale, an example item, as well as the mean, standard deviation, and reliability statistic (Cronbach's alpha) for the pooled sample (treatment and control) of data collected at baseline.

Table C.3. Baseline Social and Emotional Outcome Scale Reliability (Research Question 4)

Noncognitive Outcome	Number of Items	Example Item	Number Reporting	Mean	Standard Deviation	Cronbach's Alpha
Perception of support from adults at school[50]	15	At school, there is a teacher or some other adult who: Really cares about me.	318	5.10	1.22	0.93
Perception of peer acceptance[51]	7	Students in my classes are willing to listen to me.	334	4.38	1.34	0.89
Academic self-concept[52]	7	I am a quick learner.	339	4.59	1.04	0.75
Academic self-efficacy[53]	6	How sure are you that you can: Get to class on time?	334	5.54	1.05	0.80
Self-efficacy in help-seeking[54]	5	How sure are you that you can: Realize when you need help with something?	339	5.17	1.30	0.85
Self-efficacy in goal-setting[55]	4	How sure are you that you can: Set goals that you want to achieve?	337	5.44	1.32	0.90
Decision-making skills[56]	6	How often would you say that you: Stop to think about your choices before you make a decision?	335	4.97	1.20	0.87

All outcomes were operationalized as mean scale scores from questionnaire items with 7-point Likert-type scales. Scale scores were constructed by estimating the mean of all items that made up the scale and were only estimated if a student responded to all items within a specified scale. We did not impute any missing values in these outcome measures.

PRG underwent a lengthy instrument development process prior to the start of the first year of implementation of the AM intervention. We developed a preliminary instrument based on the program logic model and the mediating outcomes specified within. To develop this initial draft, we reviewed the published literature to find examples of instruments that measured each mediator specified and began a process of narrowing down potential scales based on the specific skills and attitudes the program reinforces. Prior to selecting a final set of items for each scale, we first solicited expert feedback from three high school teachers who work with students in the target population. During these feedback sessions, we received feedback from teachers on readability, comprehension, and relevance of items and instructions. We then conducted a pilot test where we collected survey responses from 26 ninth-grade students in the target population, but who would not be eligible for the study because they would have progressed beyond 10th grade before the first year of study implementation. Using the pilot test data, we reviewed the distribution of responses to confirm that items and response scales were

[50] Items taken from the California Healthy Kids Survey (Hanson & Kim, 2007), *school assets subscale*, the Student Engagement Instrument (Appleton et al., 2006), the Psychological Sense of School Membership Scale (Goodenow, 1993), and the Beginning High School Survey (Stoker et al., 2017).

[51] Items taken from the School Success Profile, *social support subscale* (Bowen et al., 2003), Beginning High School Survey, *peer fit subscale* (Stoker et al. 2017), and the Student Engagement Instrument (Appleton et al., 2006).

[52] Items adapted from the *Children's Motivation Toward Science scale* (Bathgate et al., 2014).

[53] Items adapted from the Beginning High School Survey, *self-efficacy subscale* (Stoker et al., 2017).

[54] Items adapted from the California Healthy Kids Survey, *problem-solving subscale* (Hanson & Kim, 2007).

[55] Items adapted from the *Goal-Setting Self-Efficacy scale* (McNeal & Hansen, 1999).

[56] Items from the *Decision-Making Skills scale* (McNeal & Hansen, 1999).

appropriately scaled. Prior to conducting any impact analyses on outcome (post-program) data, we conducted scale reliability and factor analyses on the baseline survey data to confirm final outcome scale operationalization.

ANALYTIC APPROACH

MODEL SPECIFICATIONS

As detailed in our research questions, our proposed impact study investigated whether offering the AM intervention to participants impacts school attendance, progressing in school, and student discipline at the end of participants' 10th- and 11th-grade years. We did this within the ITT framework, which does not take into account participants' actual or measured exposure to the treatment itself, but, rather, the effect of the offer of the treatment (AM) relative to the offer of receiving the control condition (class as usual). This framework maintains the integrity of the experimental structure by including all participants who were randomized (except those who attrite) in the analytic sample, maintaining an exogenous assignment of participants to the experimental condition. Under this structure, we are able to produce an unbiased estimate of the treatment effect regardless of variation in program exposure.

The benchmark analyses pooled data across study cohorts for all schools therein and estimated effects using student-level data. Specifically, the analytic sample for research questions assessed at 10th grade included all students randomized at the beginning of the school year in all three cohorts (fall 2021, 2022, and 2023) and who provided outcome data at the end of 10th grade. Analytic samples for research questions assessed at the end of 11th grade include all students randomized at the beginning of the first two cohorts (fall 2021 and 2022) and who provided outcome data at the end of 10th and 11th grades.

We used a regression-estimated approach that modeled intervention effects while controlling for relevant covariates (detailed in Table C.1). We used a model-based approach rather than a straight difference-of-means approach in order to increase the precision of those estimates. The empirical model was estimated with an Ordinary Least Squares (OLS) regression model (using Stata).[57] We model each outcome separately using the following empirical model:

$$Y_{Post} = \beta_0 + \beta_1 T + \sum(\beta_P X_P) + \varepsilon$$

where:

Y_{Post} – The outcome variable reported for each participant at the end of 10th grade (first time point) or 11th grade (second time point).

T – A dummy treatment indicator variable whose value equals 1 if the participant is randomized into the treatment group and 0 otherwise.

X – A vector of p covariates, including both baseline (i.e., measured prior to receiving intervention or exogenous to treatment) participant-level measures as well as blocking variables (i.e., school, by cohort) to account for the variation in outcomes associated with these variables and to increase the precision of our impact estimates. These covariates include: (1) the pre-intervention measure of the outcome measure; (2) age; (3) race/ethnicity; (4) gender; (5) ELL status; (6) IEP; and (7) randomization blocks. All covariates were centered at the grand mean.

[57] We conducted a sensitivity analysis that modeled the dichotomous suspension outcome using a logistic regression model.

β_0 – The intercept term, which represents the mean outcome score for control participants at the end of 10th grade (first time point) or 11th grade (second time point), with all other variables in the model held constant at zero.

β_1 – This is the parameter estimate of substantive interest. β_1 represents the adjusted mean difference in treatment and control participants' outcome score at the end of 10th grade (first time point) or 11th grade (second time point), controlling for all other variables included in the model. We report the model-estimated difference between the treatment and control group (β_1), along with the model estimates for the treatment mean ($\beta_1 + \beta_0$) and control mean (β_0). Statistical significance was based on test statistics produced by Stata for the coefficient β_1 using a two-tailed test, with $p < .05$.

The same analytic strategy was used to examine outcomes at the end of 10th and 11th grades.

TREATMENT OF MISSING DATA

We did not impute missing outcome data. Impact analytic samples included only those observations that had non-missing post-intervention data.

Missing baseline and covariate data were handled according to the techniques outlined by the National Center for Education Evaluation.[58] Missing covariate data, including baseline measures of the outcome, were treated using dummy variable adjustment according to guidance provided by Puma et al. (2009; for details, see pp. 34–35). The justification for this is that (1) our first priority is to reduce selection bias by retaining the sample that is most representative of our ITT sample; (2) covariate data are included only to increase the precision of our impact estimates; and (3) assuming low differential attrition, this should not bias results.

CALCULATION OF EFFECT SIZE

We calculate effect sizes in accordance with the guidelines published in the *What Works Clearinghouse (WWC) Procedures and Standards Handbook*, *Version 5.0*.

CONTINUOUS VARIABLES

For each of the outcomes, the standard deviation for each condition is estimated from the sample data. We calculate the pooled standard deviation using the following formula:

$$S_p = \sqrt{\frac{(n_t - 1)S_t^2 + (n_c - 1)S_c^2}{(n_t + n_c - 2)}}$$

where n_t and n_c are the sample sizes, and S_t and S_c are the student-level standard deviations for the analytic treatment and control groups, respectively.

[58] See footnote 48.

For each outcome, the standardized effect size, known as Hedges' g, is calculated using the following formula:

$$g = \frac{\beta_1}{S_p}$$

where β_1 is the regression coefficient for the intervention's effect (adjusted mean difference in the outcome variable between the treatment and comparison group), and S_p is the pooled standard deviation.

DICHOTOMOUS VARIABLES

According to the What Works Clearinghouse, "The effect size measure of choice for dichotomous outcomes is the Cox Index, which yields effect size values similar to the values of Hedges' g that one would obtain if group means, standard deviations, and sample sizes were available." Following this guidance, we used the Cox Index to estimate baseline equivalence for dichotomous baseline covariates using the following formula:

$$d_{Cox} = \left[ln\left(\frac{p_t}{1-p_t}\right) - ln\left(\frac{p_c}{1-p_c}\right) \right] \Big/ 1.65$$

where p_t and p_c represent the probability of occurrence of the event (or characteristic) within the treatment and comparison groups, respectively.

BASELINE EQUIVALENCE

Baseline equivalence of the treatment and control samples was established using baseline measures of the outcome for each analytic sample, in accordance with the *WWC Procedures and Standards Handbook, Version 5.0*. The WWC specifies that differences less than or equal to 0.05 standard deviations require no statistical adjustment for groups to be considered equivalent. For differences between 0.05 and 0.25 standard deviations, an analysis must include an acceptable statistical adjustment for the baseline characteristic to meet equivalence standards. Differences above 0.25 standard deviations in value indicate nonequivalence of groups on that baseline characteristic.

We assessed baseline equivalence of treatment and control groups within each analytic sample by assessing the pre-intervention differences in important background characteristics and outcomes observed in data. To assess equivalence, we generated a model-based estimate of the difference between treatment and control groups for the pre-intervention variables; the empirical model is a reduced form of the model used to estimate program impact (as specified in the Model Specifications section). It is a reduced form because individual-level covariates are omitted. Separate models are run, and estimates provided, for each of the variables selected for baseline equivalence. Where the baseline variable is continuous, the model is estimated with OLS and the standardized difference is calculated using the Hedges' g formula; where the baseline variable is dichotomous, the model is estimated with a logistic regression model and the difference in the probability of the occurrence is calculated with the Cox Index formula.

APPENDIX D. SUBGROUP RESULTS

In this appendix, we provide the details of subgroup analyses described in the Impact Study Results of the *Final Evaluation Report*. For each outcome, we looked at groups of students based on their race/ethnicity, academic disadvantage (English Language Learner [ELL] or Individualized Educational Plan [IEP]), students who self-reported two or more barriers to school participation, students attending rural or urban/suburban schools, and individual program eligibility criteria met to identify the potential for heterogenous program effects. We omit results from students with IEPs or with an ELL designation because of small sample sizes.

TENTH GRADE

BEHAVIORAL AND ACADEMIC OUTCOMES

Table D.1. Attendance at the End of 10th Grade

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**309**	**147.91**	**146.19**	**1.72 (2.19)**	**0.432**	**0.06**
Students identified as Hispanic/Latino/a	73	160.31	158.17	2.14 (3.05)	0.485	0.11
Students identified as Black/African American	134	148.07	148.50	–0.43 (2.93)	0.884	–0.02
Students who reported 2+ barriers to school	189	148.33	147.12	1.21 (2.98)	0.685	0.04
Rural students	197	145.73	146.86	–1.12 (2.91)	0.699	–0.03
Urban students	112	159.41	154.20	5.21 (2.91)	0.077~	0.32
Eligible because of attendance	166	142.24	144.78	–2.55 (3.43)	0.460	–0.08
Eligible because of course failure	198	145.64	144.56	1.18 (2.74)	0.667	0.05
Eligible because of discipline infractions	49	139.49	147.61	–8.12 (6.33)	0.210	–0.29

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

Table D.2. *Credits Earned at the End of 10th Grade*

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**308**	**8.03**	**7.37**	**0.67 (0.51)**	**0.196**	**0.10**
Students identified as Hispanic/Latino/a	73	9.70	8.96	0.74 (1.57)	0.641	0.09
Students identified as Black/African American	133	7.87	7.80	0.07 (0.81)	0.932	0.01
Students who reported 2+ barriers to school	188	8.44	7.70	0.74 (0.71)	0.296	0.11
Rural students	196	7.16	7.61	–0.44 (0.33)	0.179	–0.16
Urban students	112	10.14	7.68	2.46 (1.38)	0.077~	0.25
Eligible because of attendance	166	5.72	5.71	0.00 (0.50)	0.993	0.00
Eligible because of course failure	198	7.27	6.87	0.40 (0.56)	0.474	0.10
Eligible because of discipline infractions	49	–10.62	–10.39	–0.23 (1.62)	0.889	–0.03

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

Table D.3. *Suspended in 10th Grade*

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size[59]
Full sample	**302**	**0.20**	**0.30**	**–0.10 (0.05)**	**0.036***	**–0.44**
Students identified as Hispanic/Latino/a	73	0.09	0.14	–0.05 (0.07)	0.542	–0.95
Students identified as Black/African American	131	0.10	0.16	–0.06 (0.08)	0.441	–0.17
Students who reported 2+ barriers to school	184	0.16	0.27	–0.11 (0.06)	0.074~	–0.54
Rural students	190	0.26	0.29	–0.03 (0.06)	0.620	–0.14
Urban students	112	0.10	0.30	–0.20 (0.07)	0.008**	–1.15
Eligible because of attendance	160	0.12	0.22	–0.10 (0.08)	0.201	–0.47
Eligible because of course failure	194	0.32	0.42	–0.11 (0.06)	0.096~	–0.49
Eligible because of discipline infractions	47	0.78	0.93	–0.15 (0.19)	0.440	–0.62

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

[59] Effect sizes represent the Cox Index in Table D.3.

SOCIAL AND EMOTIONAL OUTCOMES

Table D.4. Perceived Adult Support at the End of 10th Grade

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**263**	**5.62**	**5.15**	**0.47 (0.16)**	**0.003****	**0.37**
Students identified as Hispanic/Latino/a	66	5.68	5.34	0.34 (0.24)	0.167	0.33
Students identified as Black/African American	116	5.47	5.11	0.36 (0.27)	0.188	0.26
Students who reported 2+ barriers to school	165	5.52	5.02	0.50 (0.20)	0.016*	0.39
Rural students	180	5.39	4.75	0.64 (0.18)	0.001**	0.51
Urban students	83	5.42	5.48	–0.05 (0.30)	0.855	–0.04
Eligible because of attendance	131	5.52	5.04	0.48 (0.22)	0.035*	0.36
Eligible because of course failure	168	5.53	4.80	0.73 (0.21)	0.001**	0.52
Eligible because of discipline infractions	47	4.50	4.29	0.21 (0.45)	0.646	0.15

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

Table D.5. Perceived Peer Acceptance at the End of 10th Grade

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**282**	**4.77**	**4.69**	**0.08 (0.13)**	**0.561**	**0.06**
Students identified as Hispanic/Latino/a	73	5.03	4.76	0.28 (0.31)	0.377	0.22
Students identified as Black/African American	124	4.40	4.38	0.02 (0.20)	0.901	0.02
Students who reported 2+ barriers to school	172	4.63	4.71	–0.08 (0.18)	0.642	–0.07
Rural students	193	4.56	4.50	0.06 (0.16)	0.686	0.05
Urban students	89	5.42	5.34	0.08 (0.28)	0.775	0.06
Eligible because of attendance	145	4.86	4.82	0.03 (0.19)	0.868	0.02
Eligible because of course failure	181	4.83	4.67	0.16 (0.18)	0.377	0.12
Eligible because of discipline infractions	50	4.01	4.32	–0.31 (0.44)	0.491	–0.25

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

Table D.6. Academic Self-Concept at the End of 10th Grade

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**282**	**4.72**	**4.59**	**0.13 (0.09)**	**0.144**	**0.13**
Students identified as Hispanic/Latino/a	72	4.73	4.77	–0.04 (0.23)	0.871	–0.04
Students identified as Black/African American	123	4.73	4.62	0.11 (0.14)	0.448	0.12
Students who reported 2+ barriers to school	173	4.68	4.53	0.15 (0.12)	0.213	0.15
Rural students	192	4.59	4.42	0.17 (0.10)	0.098~	0.17
Urban students	90	4.45	4.46	–0.01 (0.17)	0.965	–0.01
Eligible because of attendance	146	4.73	4.51	0.22 (0.13)	0.087~	0.21
Eligible because of course failure	179	4.68	4.43	0.25 (0.12)	0.038*	0.26
Eligible because of discipline infractions	50	4.34	4.42	–0.07 (0.33)	0.826	–0.07

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

Table D.7. Academic Self-Efficacy at the End of 10th Grade

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**287**	**5.66**	**5.34**	**0.31 (0.13)**	**0.013***	**0.26**
Students identified as Hispanic/Latino/a	74	5.60	5.30	0.30 (0.23)	0.205	0.29
Students identified as Black/African American	128	5.46	5.18	0.29 (0.22)	0.203	0.23
Students who reported 2+ barriers to school	176	5.54	5.43	0.11 (0.14)	0.425	0.11
Rural students	191	5.51	5.19	0.33 (0.15)	0.031*	0.28
Urban students	96	5.68	5.43	0.25 (0.28)	0.384	0.19
Eligible because of attendance	149	5.75	5.40	0.35 (0.20)	0.080~	0.29
Eligible because of course failure	181	5.48	5.03	0.45 (0.17)	0.011*	0.34
Eligible because of discipline infractions	52	5.53	5.45	0.09 (0.50)	0.862	0.07

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

Table D.8. *Self-Efficacy in Help-Seeking at the End of 10th Grade*

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**287**	**5.49**	**5.16**	**0.33 (0.16)**	**0.037***	**0.23**
Students identified as Hispanic/Latino/a	74	5.14	4.81	0.34 (0.28)	0.225	0.27
Students identified as Black/African American	127	5.59	5.44	0.15 (0.25)	0.542	0.11
Students who reported 2+ barriers to school	174	5.38	5.18	0.20 (0.24)	0.401	0.14
Rural students	194	5.09	4.76	0.33 (0.19)	0.083~	0.234
Urban students	93	5.93	5.74	0.19 (0.30)	0.534	0.13
Eligible because of attendance	146	5.68	5.37	0.31 (0.24)	0.197	0.21
Eligible because of course failure	184	5.46	5.15	0.31 (0.21)	0.149	0.21
Eligible because of discipline infractions	51	4.41	4.78	–0.37 (0.42)	0.385	–0.26

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

Table D.9. *Self-Efficacy in Goal-Setting at the End of 10th Grade*

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**287**	**5.61**	**5.32**	**0.30 (0.14)**	**0.036***	**0.22**
Students identified as Hispanic/Latino/a	73	5.60	5.32	0.28 (0.22)	0.215	0.22
Students identified as Black/African American	128	5.69	5.44	0.26 (0.22)	0.247	0.19
Students who reported 2+ barriers to school	174	5.55	5.22	0.33 (0.19)	0.076	0.26
Rural students	193	5.40	5.13	0.28 (0.17)	0.099~	0.20
Urban students	94	6.33	6.06	0.27 (0.32)	0.394	0.20
Eligible because of attendance	145	5.76	5.41	0.35 (0.22)	0.111	0.24
Eligible because of course failure	182	5.42	5.02	0.40 (0.18)	0.028*	0.27
Eligible because of discipline infractions	51	5.08	5.32	–0.24 (0.36)	0.504	–0.18

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

Table D.10. Decision-Making Skills at the End of 10th Grade

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**281**	**5.24**	**5.02**	**0.22 (0.13)**	**0.087~**	**0.18**
Students identified as Hispanic/Latino/a	73	5.05	5.14	–0.09 (0.27)	0.738	–0.08
Students identified as Black/African American	122	5.47	5.31	0.16 (0.19)	0.426	0.13
Students who reported 2+ barriers to school	171	5.20	5.04	0.15 (0.17)	0.363	0.14
Rural students	189	4.86	4.63	0.24 (0.16)	0.133	0.19
Urban students	92	5.46	5.53	–0.07 (0.24)	0.764	–0.06
Eligible because of attendance	145	5.34	5.04	0.30 (0.19)	0.116	0.25
Eligible because of course failure	178	5.22	4.95	0.27 (0.17)	0.120	0.21
Eligible because of discipline infractions	49	4.30	4.59	–0.29 (0.42)	0.491	–0.24

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

ELEVENTH GRADE

BEHAVIORAL AND ACADEMIC OUTCOMES

Table D.11. Attendance at the End of 11th Grade

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**188**	**294.02**	**303.46**	**–9.44 (6.77)**	**0.165**	**–0.17**
Students identified as Hispanic/Latino/a	37	330.51	321.79	8.73 (10.73)	0.425	0.243
Students identified as Black/African American	81	277.97	301.09	–23.12 (9.65)	0.019*	–0.49
Students who reported 2+ barriers to school	115	259.27	273.46	–14.19 (9.67)	0.146	–0.24
Rural students	143	287.98	306.38	–18.40 (7.24)	0.012*	–0.32
Urban students	45	304.89	294.77	10.12 (12.84)	0.437	0.24
Eligible because of attendance	116	286.09	306.79	–20.70 (10.94)	0.062~	–0.34
Eligible because of course failure	128	286.56	297.45	–10.89 (9.17)	0.237	–0.21
Eligible because of discipline infractions	27	289.66	320.07	–30.41 (24.68)	0.246	–0.44

Note: ~ $p < 0.10$; * $p < 0.05$; * *$p < 0.01$.

Table D.12. Credits Earned at the End of 11th Grade

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**185**	**25.05**	**24.22**	**0.83 (1.06)**	**0.432**	**0.06**
Students identified as Hispanic/Latino/a	37	35.23	34.54	0.69 (2.38)	0.777	0.04
Students identified as Black/African American	80	18.00	17.52	0.48 (1.93)	0.803	0.03
Students who reported 2+ barriers to school	114	27.58	26.28	1.30 (1.66)	0.437	0.09
Rural students	141	14.21	15.23	–1.02 (0.68)	0.136	–0.24
Urban students	44	31.54	24.35	7.18 (3.95)	0.079~	0.29
Eligible because of attendance	113	21.13	21.17	–0.04 (1.12)	0.972	0.00
Eligible because of course failure	125	17.78	17.44	0.34 (0.96)	0.727	0.04
Eligible because of discipline infractions	26	33.79	30.27	3.52 (5.41)	0.532	0.15

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

Table D.13. Suspended in 10th or 11th Grade

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size[60]
Full sample	**176**	**0.36**	**0.34**	**0.02 (0.07)**	**0.757**	**0.05**
Students identified as Hispanic/Latino/a	36	0.13	0.09	0.05 (0.13)	0.728	0.05
Students identified as Black/African American	79	0.25	0.17	0.07 (0.11)	0.515	0.29
Students who reported 2+ barriers to school	105	0.77	0.70	0.07 (0.10)	0.506	0.10
Rural students	132	0.37	0.30	0.07 (0.08)	0.402	0.26
Urban students	44	0.12	0.17	–0.05 (0.13)	0.690	–0.40
Eligible because of attendance	105	0.39	0.33	0.06 (0.10)	0.552	0.25
Eligible because of course failure	119	0.66	0.65	0.02 (0.09)	0.867	0.11
Eligible because of discipline infractions	24	1.07	1.50	–0.43 (0.30)	0.190	–1.09

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

[60] Effect sizes represent the Cox Index estimate in Table D.13.

SOCIAL AND EMOTIONAL OUTCOMES

Table D.14. Perceived Adult Support at the End of 11th Grade

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**178**	**5.96**	**5.48**	**0.48 (0.23)**	**0.042***	**0.35**
Students identified as Hispanic/Latino/a	45	5.77	6.33	–0.56 (0.66)	0.407	–0.38
Students identified as Black/African American	79	5.93	5.24	0.69 (0.35)	0.052~	0.51
Students who reported 2+ barriers to school	113	6.03	5.70	0.33 (0.24)	0.179	0.26
Rural students	142	5.53	5.01	0.52 (0.27)	0.053~	0.39
Urban students	36	6.12	5.95	0.17 (0.72)	0.819	0.12
Eligible because of attendance	106	5.84	5.78	0.06 (0.33)	0.869	0.04
Eligible because of course failure	121	6.51	5.79	0.72 (0.27)	0.010*	0.54
Eligible because of discipline infractions	26	1.83	2.19	–0.37 (0.79)	0.656	–0.21

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

Table D.15. Perceived Peer Acceptance at the End of 11th Grade

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**182**	**5.24**	**5.03**	**0.21 (0.21)**	**0.308**	**0.15**
Students identified as Hispanic/Latino/a	43	5.11	5.64	–0.54 (0.57)	0.356	–0.40
Students identified as Black/African American	82	5.27	5.13	0.14 (0.28)	0.627	0.10
Students who reported 2+ barriers to school	115	5.34	5.03	0.31 (0.24)	0.201	0.22
Rural students	146	4.68	4.36	0.31 (0.24)	0.196	0.21
Urban students	36	5.56	5.92	–0.35 (0.35)	0.316	–0.32
Eligible because of attendance	109	4.73	4.75	–0.02 (0.31)	0.948	–0.01
Eligible because of course failure	127	5.17	4.73	0.44 (0.23)	0.062~	0.31
Eligible because of discipline infractions	28	2.55	3.25	–0.69 (0.47)	0.170	–0.42

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

Table D.16. Academic Self-Concept at the End of 11th Grade

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**180**	**4.77**	**4.48**	**0.29 (0.12)**	**0.019***	**0.29**
Students identified as Hispanic/Latino/a	44	4.27	4.30	–0.03 (0.30)	0.911	–0.03
Students identified as Black/African American	79	4.56	4.28	0.28 (0.19)	0.152	0.32
Students who reported 2+ barriers to school	114	4.91	4.66	0.25 (0.17)	0.145	0.23
Rural students	145	4.66	4.38	0.28 (0.13)	0.035*	0.28
Urban students	35	5.02	4.90	0.11 (0.50)	0.824	0.12
Eligible because of attendance	107	4.78	4.41	0.37 (0.17)	0.030*	0.34
Eligible because of course failure	125	4.79	4.41	0.38 (0.16)	0.018	0.39
Eligible because of discipline infractions	28	3.60	4.16	–0.56 (0.64)	0.395	–0.48

Note: ~ $p < 0.10$; *$p < 0.05$; **$p < 0.01$.

Table D.17. Academic Self-Efficacy at the End of 11th Grade

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**185**	**5.66**	**5.56**	**0.10 (0.16)**	**0.535**	**0.09**
Students identified as Hispanic/Latino/a	47	5.76	5.46	0.29 (0.40)	0.469	0.26
Students identified as Black/African American	81	5.20	5.10	0.10 (0.26)	0.709	0.09
Students who reported 2+ barriers to school	117	5.38	5.33	0.04 (0.21)	0.841	0.04
Rural students	150	5.77	5.63	0.13 (0.17)	0.440	0.12
Urban students	35	5.97	6.23	–0.26 (0.53)	0.629	–0.26
Eligible because of attendance	110	5.62	5.48	0.14 (0.23)	0.558	0.13
Eligible because of course failure	128	5.48	5.50	–0.02 (0.19)	0.912	–0.02
Eligible because of discipline infractions	29	5.77	6.21	–0.45 (1.03)	0.673	–0.32

Note: ~ $p < 0.10$; *$p < 0.05$; **$p < 0.01$.

Table D.18. Self-Efficacy in Help-Seeking at the End of 11th Grade

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**183**	**5.35**	**5.27**	**0.08 (0.21)**	**0.704**	**0.06**
Students identified as Hispanic/Latino/a	45	5.63	5.95	–0.32 (0.58)	0.583	–0.21
Students identified as Black/African American	81	5.36	5.29	0.08 (0.29)	0.792	0.06
Students who reported 2+ barriers to school	117	5.74	5.56	0.18 (0.28)	0.516	0.13
Rural students	148	4.98	4.90	0.08 (0.24)	0.730	0.06
Urban students	35	6.59	6.70	–0.12 (0.59)	0.848	–0.10
Eligible because of attendance	109	6.18	5.86	0.32 (0.36)	0.383	0.22
Eligible because of course failure	129	5.61	5.38	0.23 (0.25)	0.361	0.17
Eligible because of discipline infractions	28	2.59	3.21	–0.61 (0.83)	0.475	–0.43

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

Table D.19. Self-Efficacy in Goal-Setting at the End of 11th Grade

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**180**	**5.65**	**5.65**	**0.00 (0.18)**	**0.980**	**0.00**
Students identified as Hispanic/Latino/a	44	5.58	5.35	0.24 (0.53)	0.660	0.17
Students identified as Black/African American	79	5.37	5.39	–0.02 (0.26)	0.934	–0.02
Students who reported 2+ barriers to school	113	5.66	5.60	0.06 (0.26)	0.812	0.05
Rural students	145	5.62	5.61	0.01 (0.20)	0.968	0.01
Urban students	35	5.49	5.61	–0.11 (0.54)	0.837	–0.10
Eligible because of attendance	107	5.36	5.18	0.18 (0.24)	0.446	0.14
Eligible because of course failure	126	5.55	5.69	–0.14 (0.22)	0.528	–0.11
Eligible because of discipline infractions	29	5.14	4.93	0.21 (0.85)	0.806	0.14

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

Table D.20. Decision-Making Skills at the End of 11th Grade

Sample	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Full sample	**183**	**5.35**	**4.92**	**0.43 (0.19)**	**0.022***	**0.35**
Students identified as Hispanic/Latino/a	45	4.95	4.42	0.53 (0.41)	0.203	0.44
Students identified as Black/African American	82	4.93	4.67	0.26 (0.27)	0.327	0.22
Students who reported 2+ barriers to school	115	5.23	4.85	0.38 (0.24)	0.112	0.31
Rural students	147	5.28	4.85	0.43 (0.21)	0.045*	0.34
Urban students	36	5.19	4.78	0.41 (0.54)	0.458	0.39
Eligible because of attendance	108	5.29	4.88	0.41 (0.27)	0.142	0.32
Eligible because of course failure	128	5.24	4.82	0.43 (0.24)	0.074~	0.33
Eligible because of discipline infractions	29	4.59	5.02	–0.44 (0.69)	0.537	–0.31

Note: ~ $p < 0.10$; * $p < 0.05$; ** $p < 0.01$.

APPENDIX E. SENSITIVITY STUDIES

In this appendix, we provide the results of sensitivity analyses conducted on each outcome for Research Questions 1 through 3. For each research question at each time point, we provide the results of alternative operationalizations of the outcome measure (i.e., attendance rate, promotion to next grade, and number of days suspended), as well as alternative analytic samples that do not exclude students with partial enrollment data (i.e., all students with any outcome data reported, regardless of whether it is a partial measure of the outcome). For Research Question 3, we also provide the results of the logic regression model on the dichotomous outcome. These details are provided in Tables E.1 through E.3. Table E.4 provides the model estimates from the exploratory analyses conducted examining 10th-grade outcomes for the subsample of our study population that contributed data at 11th grade (and which is presented in Figure 7 in the Discussion section).

Table E.1. Sensitivity Analyses: Research Question 1

Model	Number Reporting	Adjusted Mean Difference	Standard Error	*p*-value	Effect Size
End of 10th grade					
Benchmark model	309	1.72	2.19	0.432	0.06
Attendance rate in 10th grade	309	0.01	0.01	0.332	0.09
All students with outcome data reported, regardless of complete enrollment	322	2.78	2.56	0.280	0.09
End of 11th grade					
Benchmark model	188	–9.44	6.77	0.165	–0.17
Attendance rate in 11th grade	195	–0.02	0.02	0.401	–0.10
All students with outcome data reported, regardless of complete enrollment	203	–3.97	7.65	0.604	–0.06

Note: ~ $p < 0.10$; *$p < 0.05$; **$p < 0.01$

Table E.2. *Sensitivity Analyses: Research Question 2*

Model	Number Reporting	Adjusted Mean Difference	Standard Error	*p*-value	Effect Size
End of 10th grade					
Benchmark model	308	0.67	0.51	0.196	0.10
Promotion to next grade as outcome	325	0.07	0.04	0.123	0.15
All students with outcome data reported, regardless of complete enrollment	321	0.59	0.49	0.230	0.09
End of 11th grade					
Benchmark model	185	0.83	1.06	0.432	0.06
Promotion to next grade as outcome	217	–0.02	0.05	0.623	–0.07
All students with outcome data reported, regardless of complete enrollment	200	0.63	1.04	0.544	0.04

Note: ~ $p < 0.10$; *$p < 0.05$; **$p < 0.01$

Table E.3. *Sensitivity Analyses: Research Question 3*

Model	Number Reporting	Adjusted Mean Difference	Standard Error	*p*-value	Effect Size
End of 10th grade					
Benchmark model	302	–0.10	0.05	0.036*	–0.44
Count of days suspended as outcome	302	–0.46	0.40	0.253	–0.14
All students with outcome data reported, regardless of complete enrollment	334	–0.10	0.04	0.034*	–0.42
Logistic regression model	302	–0.72	0.33	0.027*	–0.44
End of 11th grade					
Benchmark model	176	0.02	0.07	0.757	0.05
Count of days suspended as outcome	176	2.02	1.94	0.300	0.17
All students with outcome data reported, regardless of complete enrollment	210	–0.02	0.07	0.791	–0.07
Logistic regression model	176	0.08	0.39	0.834	0.05

Note: ~ $p < 0.10$; *$p < 0.05$; **$p < 0.01$

Table E.4. *Tenth-Grade Outcome Effect Sizes for Full and Subsamples*

Outcome Measure	Number Reporting	Adjusted AM Mean	Adjusted Control Mean	Adjusted Mean Difference (SE)	*p*-value	Effect Size
Behavioral and academic						
Days attended in 10th grade	188	148.47	148.63	–0.16 (2.52)	0.951	–0.01
Credits earned in 10th grade	185	5.05	4.28	0.77 (0.73)	0.296	0.11
Suspended during 10th grade	176	26.21%	30.21%	–0.04 (0.07)	0.552	–0.19
Social and emotional						
Perceived adult support	146	5.86	5.17	0.70 (0.22)	0.002**	0.53
Perceived peer acceptance	159	4.81	4.73	0.08 (0.19)	0.654	0.06
Academic self-concept	181	4.90	4.78	0.12 (0.12)	0.346	0.12
Academic self-efficacy	162	5.86	5.40	0.46 (0.17)	0.008**	0.40
Self-efficacy in help-seeking	164	5.52	5.27	0.25 (0.24)	0.287	0.18
Self-efficacy in goal-setting	162	5.44	5.29	0.15 (0.20)	0.444	0.11
Decision-making skills	158	5.32	5.01	0.31 (0.17)	0.075~	0.26

Note: ~$p < 0.10$; *$p < 0.05$; **$p < 0.01$