



West Virginia GEAR UP Year 4 Interim Evaluation Report

May 2018

Submitted to:
West Virginia Higher
Education Policy Commission

Submitted by:
ICF External Evaluation Team

Table of Contents

I. West Virginia GEAR UP History and Context	1
1. The Student Success Societies (SSS) program	2
2. SSS Program Evaluation Design	2
3. Purpose of this Report.....	3
II. Data Sources.....	3
1. Extant Data from the WVDE.....	4
1.1 Behavioral engagement data.....	4
1.1.1 Attendance.....	4
1.1.2 Discipline referrals	4
1.2 Academic achievement data.....	5
2. Student Survey Data.....	5
2.1 Attitudes toward college-going.....	5
2.1.1 College-going self-efficacy (CGSE) and college-going outcomes expectations (CGOE)	5
2.2 Grit.....	7
3. Mentoring Exposure Data from SCRIBE	7
4. Interviews with Student Success Society Mentors.....	8
III. Methods.....	8
1. Evaluation Participants and Data Collection Methods	8
1.1 Site selection, student identification, and randomization	8
2. Student Success Society Mentor Identification	9
3. Analytic Approach.....	10
3.1 Estimating SSS program impacts on attitudes toward college-going, grit, behavioral engagement, and academic achievement	10
3.2 Measuring student attrition	11
3.3 Interviews with mentors	12
IV. Results.....	13
1. Implementation	13
1.1 Describing student participation in the SSS program.....	13
1.2 Findings from interviews with SSS mentors	14
1.2.1 Roles and responsibilities.....	14
1.2.2 Mentor training and ongoing support	15
1.2.3 Student recruitment and selection.....	16
1.2.4 Implementation	16
1.2.5 Use of curriculum and resources.....	17
1.2.6 Impact/lessons learned.....	18
2. The Program Impact Analysis	19
2.1 Attitudes toward college-going and grit	19
3. Behavioral Engagement	19

4. Academic Achievement	20
5. Further Exploratory Analysis	20
5.1 Subgroup impact analysis.....	20
5.2 Correlational analysis.....	20
V. Discussion	24
VI.Recommendations	25
VII. References	28
Appendix A: Opt-Out/Permission Forms and Instrument	A-1
Appendix B: Technical Detail	B-1
1. Attrition Analysis	B-1
2. Baseline Equivalence Analysis.....	B-1
3. Multivariate Regression Analysis	B-3
4. Descriptive Statistics of Analysis Samples.....	B-7
5. Correlational Analysis of Program Exposure Hours and Student Outcomes	B-10

I. West Virginia GEAR UP History and Context

Although West Virginia has historically ranked low among states in educational attainment, the past decade has seen several noteworthy improvements. The state's [four-year high school graduation rate](#) in 2016-2017 was 89.4%, nearly 14 points higher than the 2008-2009 rate of 75.5%. In addition, the [2017 West Virginia Higher Education Report Card](#) found that the total number of degrees and credentials awarded at the state's public postsecondary institutions increased by 23.5% from 2007 to 2016. Nonetheless, college enrollment rates among recent West Virginia high school graduates indicate room for growth. Their college-going rate did increase slightly from 54.7% to 55% between 2015 and 2016. Nationally, however, 69% of 2015 high school completers entered some type of postsecondary education the following fall, according to data from the National Center for Education Statistics (2017).

To enhance college access programming and promote postsecondary success, the West Virginia Higher Education Policy Commission (the Commission) received a federal Gaining Early Awareness and Readiness for Undergraduate Programs (GEAR UP) grant for 2014-2021. Serving 50 schools, including 23 high schools, West Virginia (WV) GEAR UP seeks to promote college awareness and provide academic support and related services to ensure that more low-income students maintain the academic momentum to enroll and succeed in postsecondary education after high school. The WV GEAR UP program provides seven years of ongoing support to all students in the class of 2020 (hereafter cohort students) who are enrolled in GEAR UP schools. These supports are designed to follow cohort students from the time they first enroll as grade 7 students (i.e., 2015) to the time they enroll in postsecondary institutions (i.e., 2021). GEAR UP also provides just-in-time support for grade 12 students each year (priority students). ICF has contracted with the Commission to conduct the federally required external evaluation of WV GEAR UP, including the sub-study that is the subject of this report.

As the current cohort of GEAR UP students entered high school in fall 2016, the Commission sought to expand mentoring options available to them, in part by launching an intensive student mentoring pilot program—the Student Success Society (SSS). Research shows that mentoring can have significant positive effects on student achievement as well as motivation. Overall, young adults from low-income and underrepresented groups who have a mentor are 55% more likely to enroll in college than similar students without a mentor (Bruce & Bridgeland, 2014). Students with mentors also post higher school attendance rates and exhibit better attitudes toward school (Herrera, DuBois, & Grossman, 2013).

In order to examine the impact of the SSS mentoring program, the ICF evaluation team, in partnership with the Commission, conducted a randomized controlled trial (RCT) study with grade 9 students in three purposefully selected West Virginia high schools during the 2016-2017 school year. This report describes findings from the RCT and interviews with the participating mentors at those high schools.

1. The Student Success Societies (SSS) program

Recognizing the potential value of mentoring, the Commission launched the SSS program across its 23 high schools in fall 2016. The initiative's key objectives are to help students increase confidence and self-awareness while encouraging academic success and promoting responsibility. Other goals of the program are to help students build leadership skills, promote student engagement, and prepare students to succeed in postsecondary education. All SSS mentors participate in a training program in Charleston, WV prior to starting their work. To document student participation in the program, mentors use standard WV GEAR UP sign-in sheets for all SSS meetings and projects.

The program design stipulated that mentors meet with a group of 10-12 students every week to cover a curriculum of 10 lessons on topics such as High School Success, Grit, Learning Styles, Emotional Intelligence, Social Competencies, and Academic Success. Within each lesson is a deep-dive session that allows society members to dig deeper into a topic through videos, activities, and talking points. The curriculum also contains two booster sessions, one focused on college readiness in the fall and another on career readiness in spring of the academic year. These booster sessions coincide with key points in the annual GEAR UP calendar: College Application and Exploration Week in November and Career Exploration Month in April. Mentors are also asked to work with students to design a group or growth project that supports the overall goals and mission of WV GEAR UP. In addition, a growth project should demonstrate what students have learned during the year and provide an opportunity to reflect on their SSS experience.

While the initiative operated across most WV GEAR UP high schools, this report focuses solely on the three high schools selected for the RCT to assess the impact of the SSS program.



2. SSS Program Evaluation Design

The purpose of conducting an RCT study was to gather high-quality causal evidence about the benefits of providing intensive mentoring support—via the SSS mentoring program—to students in three high-need WV high schools. The study examined administrative and survey data from the 2015-2016 (baseline) and 2016-2017 (outcome) school years. Within each school, ICF and the Commission devised a process that randomly assigned students to either receive the offer of participating in the SSS mentoring program (treatment group) or to receive business-as-usual support through WV GEAR UP (control group). Both groups had access to regular GEAR UP services. So, in effect, this study examined the added impact of mentoring services.

After the program's first year, ICF examined its impact on student attitudes about college-going, behavioral engagement, and academic achievement using a single-level, fixed effect regression model to estimate the effect of the offer of participation in the SSS program on the proposed student outcomes. Because this model estimated the effect of the offer of participation in the SSS program, students who were assigned to the treatment group but did not receive any mentoring services were included in the analyses. ICF evaluators additionally conducted exploratory analyses to look closely at relationships between participation in the SSS program and outcomes. To better understand the implementation of the SSS mentoring program, ICF

also conducted telephone interviews in April 2018 with the seven mentors who provided intervention services to students at the three RCT schools.

3. Purpose of this Report

The purpose of the WV GEAR UP Year 4 Interim Evaluation Report is to describe the impact of offering an intensive student mentoring program—the SSS program—upon student outcomes in three primary domains: (1) attitudes about college-going, (2) behavioral engagement, and (3) academic achievement. Specifically, we conducted a systematic investigation of six research questions developed by ICF in close collaboration with the Commission (See Table 1).

Table 1. Outcomes, Domains, and Associated Research Questions

	Outcome	Outcome Domain	Research Question
1a	College-going self-efficacy (CGSE)	Attitudes about college-going	<i>What is the impact of offering the SSS intensive student mentoring program upon students' college-going self-efficacy as measured by annual WV GEAR UP student surveys?</i>
1b	College-going outcomes expectations (CGOE)	Attitudes about college-going	<i>What is the impact of offering the SSS intensive student mentoring program upon college-going outcomes expectations as measured by annual WV GEAR UP student surveys?</i>
1c	Grit	Attitudes about college-going	<i>What is the impact of offering the SSS intensive student mentoring program upon student grit scale (student efficacy scale) as measured by annual WV GEAR UP student surveys?*</i>
2a	Unexcused absence rate	Behavioral engagement	<i>What is the impact of offering the SSS intensive student mentoring program upon students' behavioral engagement as measured by total and unexcused absences, and number of discipline referrals?</i>
2b	Disciplinary referrals	Behavioral engagement	<i>What is the impact of offering the SSS intensive student mentoring program upon students' behavioral engagement as measured by total and unexcused absences, and number of discipline referrals?</i>
3	Grade point average (GPA)	Academic achievement	<i>What is the impact of offering the SSS intensive student mentoring program upon students' academic achievement, as measured by grade-point averages?*</i>

*The analyses of grit and GPA outcomes are considered exploratory due to the fact that ICF only had access to post-intervention data. Thus, for the associated research questions, we suggest careful and conservative interpretation of results. Even if the program impacts were positive and statistically significant, future replication with pre-intervention data would be necessary to draw definitive conclusions.

A secondary purpose of this report is to describe implementation of the SSS mentoring program at the three sites participating in the RCT. Understanding the implementation of the SSS program helps to contextualize the results of the RCT.

II. Data Sources

This study was designed as a low-cost, highly rigorous examination of program outcomes. To control costs and reduce burden upon participants, the majority of the data gathered for this

study came from pre-existing student-level records housed in the West Virginia Educational Information System (WVEIS). WVEIS is a statewide administrative data system maintained by the West Virginia Department of Education (WVDE) and used by all 55 counties. ICF executed a series of research agreements with each county and the WVDE in order to receive these data. As explained below, WVEIS provided student roster and demographic information as well as outcome data measuring the domains of behavioral engagement and academic achievement.

ICF gathered information about student attitudes toward college-going from the online surveys administered as part of ICF's broader WV GEAR UP evaluation. Pre-intervention surveys were administered as part of the Year 3 evaluation, while post-intervention data came from the Year 4 evaluation. These surveys provided data on three student outcomes, described in Table 1 (i.e., CGOE, CGSE, and grit).

The evaluation team gathered additional qualitative information from individual interviews with mentors in the three RCT high schools. Working with the Commission, ICF devised an interview protocol to ask these mentors about their training, their implementation of the curriculum, and their perceptions of program implementation and impact. Evaluation team members conducted these interviews individually by telephone in April 2018. The team audio-recorded all calls and analyzed transcripts with an emphasis on successes and challenges as well as variation in implementation of the curriculum across sites.

1. Extant Data from the WVDE

As noted above, ICF gathered the majority of the data required for this study from the WVDE. Outcomes of interest included student attendance, discipline referrals, and GPA. The WVDE also provided basic student-level predictors necessary for the analysis models such as gender, race and ethnicity, special education status, and pre-intervention standardized mathematics test scores. ICF requested data for the 2015-2016 (pre-intervention) and 2016-2017 (intervention) school years.

1.1 Behavioral engagement data

1.1.1 Attendance

The WVDE provided several student attendance variables from WVEIS, from which ICF constructed an unexcused absence rate for each student for each school year. The team created this outcome by dividing the number of unexcused absences accumulated during the school year by the total number of days each student was enrolled at the school (i.e., membership days).

1.1.2 Discipline referrals

When students violate school rules, the state requires that schools report such incidents through WVEIS. The WVDE provided ICF with the total number of discipline referrals for each student during school years 2015-2016 and 2016-2017. Notably, these figures included only incidents in which the student was identified as the aggressor, not those incidents in which they were reported as the victim. For this study, the total annual count of discipline referrals per student was defined as an outcome variable.



1.2 Academic achievement data

All WV public school students are required to take standardized state assessments in multiple subject areas at certain grade levels. The state policies regarding these annual assessments changed during the course of our study, posing some complications for our estimation of effects upon academic achievement outcomes. Specifically, the ninth graders who were included in our study sample took standardized mathematics and English/language arts (ELA) assessments at the end of the eighth grade—the year prior to our study. However, due to a state policy change, standardized assessments were not administered to these students again at the end of ninth grade—the end of the intervention year.

To address this issue, we sought to gather pre- and post-intervention GPA data for the students in our sample as measured at the end of eighth and ninth grades. Unfortunately, it was determined that we could only retrieve GPA data from the transactional WVEIS data environment, updated daily, which meant we could not obtain GPA data for the pre-intervention year. As a compromise, we used eighth-grade standardized test performance data as the pre-intervention measure and cumulative ninth-grade GPA data as the post-intervention measure of academic achievement. Ultimately, we chose the standardized mathematics assessment data as the single pre-intervention source in our model.¹

2. Student Survey Data

ICF obtained data from the student surveys administered as part of its broader GEAR UP evaluation to measure three outcomes in the domain of attitudes toward college-going (i.e., CGSE, CGOE, and grit). The pre-intervention survey was administered shortly after randomization was completed and the mentoring intervention began, from November 2016 through January 2017. The post-intervention survey was administered after completion of the first year of the mentoring intervention, from November 2017 through January 2018. Deriving data from this existing survey allowed ICF to gather important attitudinal outcomes with only minimal burden and no additional cost to the Commission.

2.1 Attitudes toward college-going

2.1.1 College-going self-efficacy (CGSE) and college-going outcomes expectations (CGOE)

ICF first administered CGSE and CGOE items during the Year 2 evaluation of WV GEAR UP, adapting these scales from Gibbons (2005). In short, CGSE represents an estimate of the degree to which students feel confident about their ability to prepare for and successfully get into college, while CGOE measures student perceptions of the extent to which they believe they can succeed after entering college.

¹ Our analysis showed that mathematics test scores and ELA state test scores were highly correlated (Pearson correlation .73, n=423) and the degree to which math scores and ELA scores were correlated with six outcomes were approximately the same. For example, the math scores-GPA correlation and the ELA scores-GPA correlation were, respectively, .59 and .56. Either one, therefore, can serve as a proxy of pretest achievement scores. Since math scores were slightly stronger correlates than ELA scores for a majority of six outcomes (four out of six), it was decided that math scores would be the pretest achievement proxy in all of our models.

As shown in Table 2, the CGSE scale includes items such as: “I can find a way to pay for college,” “I can get accepted to a college,” and “I can have family support for going to college.” The CGOE scale includes items such as: “I could pay for each year of college,” “I could get A’s and B’s in college,” and “I could get my family to support my wish of finishing college.” Both scales used a five-point, Likert-type response scale where 1 = *don’t know*, 2 = *not at all sure*, 3 = *somewhat sure*, 4 = *sure*, and 5 = *very sure*. The reliabilities for these measures (estimated from the collected data) were high (Cronbach Alpha [α] = 0.95 for CGSE and 0.97 for CGOE).

Table 2. Survey Items for CGSE and CGOE Scales

1. College-Going Self-Efficacy (CGSE)
How sure are you about being able to do the following?
1. I can find a way to pay for college.
2. I can get accepted to a college.
3. I can have family support for going to college.
4. I can choose a good college.
5. I can get a scholarship or grant for college.
6. I can make an educational plan that will prepare me for college.
7. I can make my family proud with my choices after high school.
8. I can choose college courses that best fit my interests.
9. I can pay for college even if my family cannot help me.
10. I can get good grades in my high school math classes.
11. I can get good grades in my high school science classes.
12. I can choose the high school classes needed to get into a good college.
13. I know enough about computers/technology to get into college.
2. College-Going Outcomes Expectations (CGOE)
If you do go to college, how sure are you about being able to do the following?
1. I could pay for each year of college.
2. I could get A’s and B’s in college.
3. I could get my family to support my wish of finishing college.
4. I could take care of myself in college.
5. I could fit in at college.
6. I could get good enough grades to get or keep a scholarship.
7. I could finish college and receive a college degree.
8. I could care for my family responsibilities while in college.
9. I could set my own schedule while in college.
10. I could make friends at college.
11. I could get the education I need for my choice of career.
12. I could get a job after I graduate from college.
13. I would like being in college.
14. I could pick the right things to study at college.
15. I could do the classwork and homework assignments in college classes.

2.2 Grit

In addition to the CGSE and CGOE scales, students completed the eight-item grit questionnaire (Duckworth & Quinn, 2009; Duckworth, Peterson, Matthews, & Kelly, 2007). The scale composed of these items measures the degree to which students are able to persevere when faced with academic challenges or adversity. For example, items on this scale included: “*New ideas and projects sometimes distract me from previous ones*” and “*Setbacks don’t discourage me.*” Table 3 shows all eight items. Response patterns are *Not like me at all*, *Not much like me*, *Somewhat like me*, *Mostly like me*, *Very much like me*. For items 2, 4, 7, and 8, the numeric values of 1 to 5 were given to these responses (e.g., *Not like me at all* = 1; *Very much like me* = 5). For items 1, 3, 5, and 6, the reversed values (i.e., 5 to 1) were given to the responses (e.g., *Not like me at all* = 5; *Very much like me* = 1). This adjustment makes all items correlate with one another positively. The reliability of the scale, estimated using Cronbach’s α (estimated from the data), was low (.50); caution is required when interpreting the results of this analysis.

Table 3. Eight-Item Grit Scale

Grit Items
1. New ideas and projects sometimes distract me from previous ones.
2. Setbacks don’t discourage me.
3. I have been obsessed with a certain idea or project for a short time, but later lost interest.
4. I am a hard worker.
5. I often set a goal but later choose to pursue a different one.
6. I have difficulty maintaining my focus on projects that take more than a few months to complete.
7. I finish whatever I begin.
8. I am diligent.

3. Mentoring Exposure Data from SCRIBE

ICF collected data measuring each student’s level of exposure to the mentoring intervention from the Standardized Collection and Reporting of Information Benefitting Education (SCRIBE) data warehouse. The Commission uses the SCRIBE system to store and report upon student, parent/guardian, and school faculty participation in all GEAR UP services. From SCRIBE, ICF evaluators exported the total number of hours of mentoring services received by each treatment and control group student during the 2016-2017 school year. We used these data to estimate total exposure to the intervention for treatment students, and to verify that control group students did not receive mentoring services.

Notably, Commission staff entered exposure data retrospectively into SCRIBE using paper and/or electronic sign-in sheets provided to the Commission by each school. At the request of the ICF evaluation team, GEAR UP regional program directors checked the exposure data for accuracy and verified its veracity prior to analysis.

4. Interviews with Student Success Society Mentors

To gather more information about the implementation of SSS programs and the school contexts in which they operated, ICF conducted telephone interviews in April 2018 with seven mentors at the three RCT schools. Since the telephone interviews were conducted during the 2017-2018 school year—the year following the RCT—not all mentors from RCT schools were interviewed. Specifically, some of the mentors who participated during the RCT year were no longer involved in the program in April 2018 and so were not interviewed. This data collection supplements the rigorous RCT and quantitative analysis by providing insight into the fidelity of program implementation at each site. The interviews provided qualitative data on topics such as the mentor training, ongoing support for mentors, implementation, and mentor perceptions of the curriculum.

III. Methods

This study utilized an RCT design to evaluate the impact of the Student Success Societies mentoring program. The study examined student outcomes in three domains: (1) attitudes toward college-going, (2) behavioral engagement, and (3) academic achievement. In designing the study, ICF followed the education evaluation research guidelines by the What Works Clearinghouse (WWC) Evidence Standards, Version 3.0 (United States Department of Education, Institute of Education Sciences, 2013).² The following section provides details about the study design and analysis methods used.

1. Evaluation Participants and Data Collection Methods

Evaluation participants included students in the three participating high schools who were offered mentoring (treatment) through the SSS program and those who were not (control). Mentors who delivered the program were also participants in the telephone interviews regarding implementation. This subsection provides additional detail regarding the evaluation participants and associated data collection methods.

1.1 Site selection, student identification, and randomization

The setting for the study consisted of three West Virginia high schools: (1) Mingo Central Comprehensive High School in Mingo County, (2) Oak Hill High School in Fayette County, and (3) Nicholas County High School in Nicholas County. These schools are a subset of the 23 public high schools currently participating in the WV GEAR UP program. The Commission purposefully selected these three schools to participate in the RCT based on three factors: (1) each school's willingness to allow random assignment as the mechanism for identifying mentoring students, (2) geographic representativeness—a single school in each of the three

² The What Works Clearinghouse (WWC) provides methodological guidelines for evaluation studies of educational programs and practices. Using the standards, WWC reports on educational intervention programs based only on well-designed and well-executed randomized control trial (RCT) or quasi-experimental design (QED) studies. For more information, please visit <https://ies.ed.gov/ncee/wwc/>.

regions served by the program was nominated by the Commission, and (3) the fact that the school had sufficient student population to support the study's sampling requirements.

After schools agreed to participate, the Commission assembled a roster of all currently enrolled ninth grade students at each school which included basic demographic information. The ICF team completed random assignment at the student level from this roster. Prior to randomization, the Commission also sent home an opt-out permission form allowing parents/guardians to request that their child not be offered mentoring services through the SSS program (see Appendix A). The Commission staff sent completed opt-out forms to ICF and removed the associated students from the sampling roster. At the request of participating schools, ICF also removed all students who had Individualized Education Plans that would be violated by the data collection procedures of the proposed study. These students were also removed from the analyses.

From the remaining subpopulation of ninth grade students, ICF selected a random sample of 150 students per school. Using a lottery system, the evaluation team randomly assigned 30 students per school to the treatment condition and 120 students per school to the control condition. Since there were more than 150 students at the participating schools, there were some students who were not assigned to either group and were not included in the analysis. All students received GEAR UP intervention services, but treatment group students also received the offer of additional mentoring through the SSS program.

These sampling targets were selected primarily for two reasons. First, since the SSS program is resource- and time-intensive, each of the three participating schools had the capacity to provide services to only 30 cohort students. Randomization was the fairest way to determine who would receive the offer and allowed for the evaluation team to reduce selection bias. Second, the number of control group students was set to three times the number of treatment students (i.e., 120 per school) to increase the study's statistical power. The resulting sample included three high schools, 90 treatment students and 360 comparison students, and the minimum detectable effect size (MDES) was estimated at .23, a small effect.

The regional program directors worked with schools to recruit mentors for participation in the SSS mentoring program during the pilot year at the RCT schools. Mentors received training by the Commission that included an overview of the curriculum as well as a presentation from a motivational speaker.

2. Student Success Society Mentor Identification

The evaluation team worked with WV GEAR UP regional program directors to identify mentors to interview in April 2018. The directors provided email and phone contact information for the seven mentors at RCT schools. ICF conducted extensive outreach to schedule the interviews, some of which were completed shortly before this report's submission. Because many WV GEAR UP schools that were not part of the RCT offered SSS programs, the evaluation team also sought to interview up to 50% of mentors at those non-RCT schools. Again working with regional program directors to reach mentors, ICF conducted interviews with eight mentors at non-RCT schools in April 2018. The evaluation team asked all mentors, whether employed at an RCT or non-RCT school, the same set of questions, with the Commission providing input for this interview protocol. All mentors consented to the audio-recording of interviews, and the

evaluation team analyzed transcripts to identify trends. The Year 4 Annual Evaluation Report in fall 2018 will analyze the interviews with mentors at non-RCT schools and note any differences in perceptions between those at RCT and non-RCT schools. This interim report focuses solely on findings from interviews with mentors at the RCT sites.

3. Analytic Approach

The ICF evaluation team used a range of analytical methods to estimate the impact of the SSS mentoring program, measure student attrition, and better understand implementation of the program at the three RCT sites.

3.1 Estimating SSS program impacts on attitudes toward college-going, grit, behavioral engagement, and academic achievement

This study examined the impact of the SSS program on student outcomes in three domains by comparing the average outcome differences between the treatment and control group students. Table 4 summarizes the outcomes. The evaluation team used statistical models to assess the group differences in outcomes after removing the influence of other predictors, such as pre-intervention test scores and student demographic variables (see Appendix B for technical details of regression analysis).

Table 4. Summary of Research Questions and Outcomes

Research Questions	Outcome
1a	College-going self-efficacy (CGSE)
1b	College-going outcomes expectations (CGOE)
1c	Grit scale
2a	Unexcused absence rate
2b	Disciplinary referrals
3	Grade point average

For the analysis of all six outcomes, ICF used Ordinary Least Squares (OLS) regression models.³ Each model estimated the size of the program impact and evaluated whether the effect size was statistically significant (alpha level = 0.05). To evaluate the standardized effect sizes, ICF relied on the WWC's use of a size larger than 0.25 as a substantively important effect size. The program impact was adjusted for other important predictors, such as pretest scores for the outcome variable, pretest mathematics test scores, gender, race and ethnicity, student status (socioeconomic level and special education), and school differences. Socioeconomic status (low and high) was determined by student status on the supplemental nutrition assistance program (SNAP). For two outcomes, grit and GPA, pretest data were not available. ICF used state math assessment test scores as the pre-intervention control for academic achievement. For grit, we

³ One important consideration in studies such as this one is the fact that students are nested within schools, and statistical models should typically account for this nesting. Due to the small number of schools (n=3), ICF decided not to use HLM to reflect this nesting. Instead, the OLS regression model included a series of binary variables (0 or 1) to indicate schools (e.g., Oak Hill = 1 if Oak Hill high school student; otherwise = 0), which adjusted for school outcome differences in the program impact estimation.

relied on pretest CGSE (college-going self-efficacy) survey findings as a baseline control because of its conceptual similarity to the meaning of grit.

3.2 Measuring student attrition

RCT designs are considered the “gold standard” in social science research because they allow researchers to draw causal conclusions about program impacts.⁴ To achieve the highest methodological rigor, RCT studies must avoid losing too many study participants in the treatment and control groups—a process known as attrition. Overall attrition, as well as unbalanced attrition rates between the two groups, can create bias in estimates of program impact. If, for example, more treatment students drop out of a study than control students, and their reason for leaving is not liking the program, the two groups will become qualitatively different from one another in student characteristics and orientation, which may bias the estimation of program impact.⁵

To address this concern for this study, ICF assessed whether school and student attrition levels were within the ranges considered acceptable by the industry standard, the WWC Evidence Standards. The study included three high schools, none of which dropped out of the study. Thus, school attrition was not a concern. As Table 5 depicts, the ICF team calculated student-level attrition based on the original student roster from the time of randomization, and the final list of students who remained in the analysis sample (after removing those who had missing values for any variables in the statistical model).

All six outcomes showed low levels of attrition based on the WWC Evidence Standards. This finding means that the data were collected according to the original plan, and that program impact estimates from these models will be considered unbiased. (See Appendix B for technical details of attrition analysis.)

⁴ RCT studies use randomization to support such conclusions. Because assignment to treatment or control conditions is completely random, the differences in student outcomes between the two groups observed at the end of the study are considered attributable only to the intervention the treatment subjects received.

⁵ Attrition can occur when participants formally drop out of a study, or when complete data records are not available for a participant. Moreover, attrition is measured separately for each outcome estimated by researchers, and separately for (1) all participants, (2) the treatment group, and (3) the control group. Although it is important to avoid high levels of overall attrition, it is most important to avoid high levels of differential attrition across the treatment and control groups. When one group leaves the study at a higher rate than the other, this can lead to biased estimates.

Table 5. Student Attrition Analysis for Six Student Outcomes

	The Whole Sample	Treatment Group	Control Group	Summary	
Outcomes	Attrition rate	Attrition rate	Attrition rate	Differential attrition	WWC attrition level
CGSE	24%	20%	26%	6%	Low
CGOE	25%	21%	26%	5%	Low
Grit	25%	20%	26%	6%	Low
Unexcused absence rate	6%	4%	6%	2%	Low
Disciplinary referral	6%	4%	6%	2%	Low
GPA	11%	11%	11%	0%	Low

Note: The ICF team used the WWC's liberal standard for rating attrition. When overall attrition is 26%, the differential attrition (the difference between the attrition rates of the treatment and control groups) must be smaller than 4.7% (if using the conservative threshold) or 9.0% (if using the liberal threshold).

3.3 Interviews with mentors

After conducting interviews with all mentors at the RCT schools, the evaluation team analyzed interview transcripts and coded results under six key themes:

- **Roles and responsibilities**, including the responsibilities of each mentor and their other roles within the school (i.e., teacher, assistant principal);
- **Mentor training and ongoing support**, which consists of their feedback on SSS training provided by the Commission and ongoing support from WV GEAR UP regional program directors and school site coordinators;
- **Student recruitment and selection**, with primary focus on the selection process at RCT sites, mentors' views of this selection process, and the commonalities/differences they perceive in characteristics of the selected students;
- **Implementation**, including the length and frequency of meetings, access to students, school-level implementation factors, and recommendations for students/mentors to make the experience successful;
- **Use of curriculum and resources**, such as the successes and challenges in implementing the formal curriculum plus mentor use of resources/web sites, one-to-one check-ins, and deep dives/group projects;
- **Impact/lessons learned**, including mentors' perceptions of the program's impact on grades, attendance, disciplinary referrals, social-emotional health, and other elements.

By aggregating comments under these themes, the evaluation team arrived at a structured analysis of the program's implementation and explored mentor training as well as mentor perceptions of the program during the RCT period. Because these interviews took place a year after the RCT period, the team also sought responses about implementation during the 2017-

2018 academic year, when sites had an opportunity to build on the foundations established during the RCT study period.

IV. Results

This section provides detailed results regarding program implementation, attitudes toward college-going and grit, behavioral engagement, academic achievement, and further exploratory analysis.

1. Implementation

Understanding program implementation at each of the three RCT sites requires both an understanding of the degree to which students in the treatment group received services related to the SSS mentoring program as well as nuances in service delivery. The implementation results described in this section stem from descriptive analysis of SSS program exposure data as well as qualitative analysis of mentor interview data.

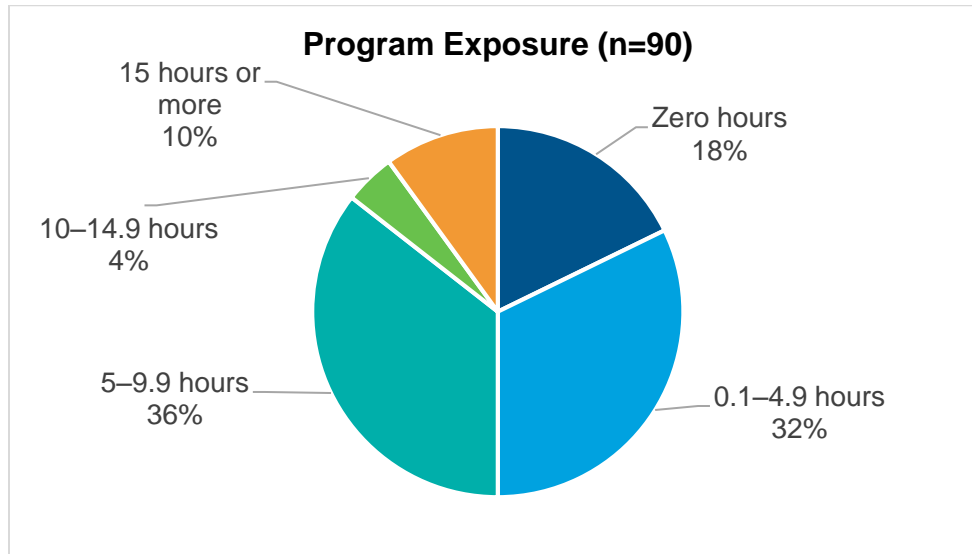
1.1 Describing student participation in the SSS program

Descriptive analysis of program exposure time (i.e., hours of mentoring received) revealed that students may not have participated in the SSS program at a sufficient level to experience positive effects.

WV GEAR UP regional program directors determined in advance that at least 10 hours of mentoring constitutes an effective dose of the program. Table 6 and Figure 1 show that 18% of treatment students received zero hours of mentoring (Group 1) and nearly one-third of the sample (Group 2) received a relatively small dose of 0.1 to 4.9 hours of mentoring. Notably, only 14% of treatment students (Groups 4 and 5 combined) met the benchmark of receiving a minimum of 10 hours of program exposure. This finding is incredibly important as it casts some doubt on whether or not the remaining study results can stand as an accurate estimation of the program's impact. Later in this report, we provide additional analysis of the relationships between program exposure and various student outcomes.

Table 6. Program Exposure Time for Treatment Students (n=90)

Subgroups by Exposure Hour	N. of Students	Percent	Cumulative Percentage
(01) Zero hours	16	18%	18%
(02) 0.1 – 4.9 hours	29	32%	50%
(03) 5 – 9.9 hours	32	36%	86%
(04) 10 – 14.9 hours	4	4%	90%
(05) 15 or more hours	9	10%	100%

Figure 1. Proportion of Treatment Students by Program Exposure Hours

Examining exposure data further, Table 7 presents average exposure hours. When all treatment students (n=90) were considered, the average program exposure time was only 5.5 hours (ranging from 3.6 hours at Nicholas County High School to 8.2 hours at Oak Hill High School). When students with zero program hours (n=16) were excluded from the analysis sample, the average program exposure time increased to 6.7 hours. Nevertheless, only 18% of that latter group of students had 10 hours or more of program exposure.

Table 7. Average Program Exposure for Treatment Students

	All Treatment Students	Treatment Students with at Least 0.1 Hours or More of Exposure
Number of students	90	74
Average exposure hours	5.5 hours	6.7 hours
% of students who received 10 hours or above	14%	18%

Note: School-specific average hours for all treatment students were 4.7 for Mingo Central High, 3.64 for Nicholas County High, and 8.2 for Oak Hill High (n=30 for schools). For treatment students with at least 0.1 hours or more of exposure, the average hours were, respectively, 5.3, 5.5, and 9.1 (n=27, 20, 27).

1.2 Findings from interviews with SSS mentors

As noted previously, the evaluation team coded the transcripts of mentor interviews and grouped their responses under six themes: Roles and Responsibilities, Mentor Training and Ongoing Support, Student Recruitment and Selection, Implementation, Use of Curriculum and Resources, and Impact/Lessons Learned. This sub-section breaks out major findings within each theme.

1.2.1 Roles and responsibilities

Of the seven mentors currently working at RCT sites, six are teachers and one is an assistant principal. Three are science teachers, while the others are instructors in music/drama, special

education, and pre-engineering. All reported extensive involvement in their schools, serving as leaders of clubs, school drama productions, the National Honor Society, and event committees. Most believed that their past record of involvement was a factor in their selection. However, most also cited an outgoing personality and commitment to youth as essential to success in the mentor role.

I believe you have to want to do this. It has to be a passion for you. You can't just see it as a job ticket or a little bit of extra money every semester.

You need to be someone the student can go to and talk if they need something.

Mentors said they spent an average of eight to 12 hours per month on mentoring activities, although this number may underestimate the informal contact that mentors have with students during a typical school week. One mentor reported that he used his planning period to pull out students for one-to-one talks (with teacher permission), and others said they used planning periods to communicate with other mentors and plan activities. The teachers were most likely to indicate that they had sufficient time to meet regularly with students, particularly if their school set aside dedicated time such as an advisory or lunch period for the SSS program.

1.2.2 Mentor training and ongoing support

All seven mentors at RCT schools found the annual training from the Commission helpful. They had extensive praise for the motivational speaker at these events, who shared his own story, generating enthusiasm among the group. According to one mentor:

Sometimes it's easy to forget the impact you can have on kids. I think his story really reminds you of that.

One mentor who attends many professional development sessions generally noted that the motivational speakers and the high-quality training offered by the Commission made for an informative day.

I can honestly say that it's one of the best trainings that I go to each year.

One mentor suggested that future trainings should include time to help new mentors acclimate to the program. For example, the Commission could take time to model the curriculum for new mentors. Another mentor suggested offering additional training or guidance about paperwork requirements, including the maintenance of logs on all contacts with students.

Through the training, mentors said that they generally came to understand their roles and the populations to be served. As one noted, the goal is to work “with those [students] who need that extra push or motivation.”

When they had questions, mentors said they received quick responses from WV GEAR UP's regional program directors, typically within an hour of calling or emailing with a question. WV GEAR UP site coordinators generally remained in close contact with mentors as well, with some attending SSS sessions or events.

1.2.3 Student recruitment and selection

Because of the study's RCT structure, this project relied on random assignment to select students for the mentoring program at all three schools during the 2016-2017 school year. To varying degrees, most mentors believed that the random assignment led to their inability to serve the neediest students. In addition, some mentors thought that when the neediest students were randomly selected to receive the offer of mentoring, they often did not stay with the program. The following statements from two mentors summarized these views:

Students randomly selected did not include all the ones who needed the program the most. Then some who needed to stay, dropped. The ones that had support at home and were already going to go to college, those kids stayed.

The ones that we needed to target are the ones who quit coming to meetings. They chose our students for us, so we didn't get to really target the ones I felt like we needed to.

However, one mentor differed from others in stating that the randomization worked fairly well in targeting high-risk youth.

Some of them probably do not need the mentoring, maybe two or three of them. The rest of them, they really do. They do need mentors. They did a pretty good job because they picked some pretty tough kids that needed mentoring.

Mentors said that for the 2017-2018 year, after the RCT period, they actively recruited new students to replace those who did not continue. They drew on teacher recommendations and their own knowledge to select students whom they believed had the greatest need for assistance.

1.2.4 Implementation

One core requirement for SSS programs is that mentors meet regularly with students to deliver a curriculum with modules in such areas as learning styles, emotional intelligence, academic success, and career planning. Based on interviews, however, two of the three sites encountered challenges carving out a regular time for students to meet with mentors during the 2016-2017 year. At Nicholas County High School, mentors and students were able to meet regularly during a lunch period, although students did not always attend. At Oak Hill and Mingo Central high schools, mentors met with students less regularly, which led to more informal contact. According to one mentor:

We just had to bring them in when we could get them, and it was really hard to get them all together at one time.

In addition, one mentor at Oak Hill dropped out soon after the start of the program in the 2016-2017 year, according to another mentor. After that departure, the school reassigned those students to the school's other two mentors.

In interviews with mentors, it became evident that Mingo Central took a different approach to organizing its students, compared with the other schools. At Mingo Central, a female mentor worked with girls while male mentors worked with the boys selected for the program. No other site employed this approach, which was not part of the original study design.

Despite challenges to the frequency of sessions, however, most mentors believed they accomplished their goal of conducting monthly one-to-one meetings with students. In some cases these meetings were informal, with mentors and students meeting for five minutes during lunch to talk. One mentor used his planning period for these meetings. Mentors also said that students were strongly engaged in this process, as were the mentors themselves. Once, a mentor found so many students struggling in the same class that he discarded his originally planned lesson to hold a group discussion about why so many students earned low grades in the class.

Mentors overwhelmingly believed that students engaged with them in these one-to-one meetings.

They'll be honest with me and tell me why they're not doing well in this class or that class, and they're open. We try to get that relationship where they can be honest.

Because of this one-on-one contact, they're able to tell me things that they're not able to say in front of everybody else.

Even in class settings, mentors said it is important to vary their approaches and remain flexible to adapt to changing circumstances. Several mentors observed that students sometimes just needed a place to talk, as typified by this comment:

I hate to call it a gripe session, but sometimes they need a place to release some frustration or just talk about a situation they're dealing with, and so I'm somebody they can come to and share concerns, whether it's college or something going on at school or a home-life situation. I'm just their bounce-off person if they need to share anything that they feel is in their heart.

1.2.5 Use of curriculum and resources

Sites differed in adherence to the curriculum, largely due to the frequency of their regular SSS meetings. Mentors at two schools cited consistent use of the curriculum, with several workshops especially popular and well-received, such as those on grit, career exploration, and emotional intelligence. One mentor noted that he had modified the career exploration lesson to include guest speakers and an activity in which students mapped out their future careers.

Mentors were most likely to use the College Foundation of West Virginia (CFWV) website, among the resources, in part because of its close connection to GEAR UP goals. One mentor reported that he adapted a goal-setting activity and the graduation pledge directly from the CFWV site.

Despite praise for the curriculum, three mentors believed that these sessions could seem too much like a class, with students becoming less engaged as a result.

It has to be fun for kids that age. It has to have some sort of fun in it. It can't just be a class, and that's kind of what it felt like to them.

I think the curriculum itself, if you don't do something to go above and beyond, can be a little boring.

One mentor from Nicholas County reported that some SSS students were also in the GEAR UP HEROs (Higher Education Readiness Officers) program, which met at the same time. They believed that many chose HEROs over SSS because the former seemed less like a traditional school-day class.

However, most mentors praised the SSS group activities, which, across the three sites, included a coat drive, a lake clean-up project, a toy drive, and an animal shelter project. Mentors believed that these projects distinguished SSS from a typical class and encouraged team building and volunteerism. For the coat and toy drives, one mentor brought students along to give the donations in person, so that they could see the impact of their volunteer efforts.

I like the group projects. We let students see how their donations and their work affected people.

1.2.6 Impact/lessons learned

Most mentors believed the SSS mentoring program is effective and needed, even if they could cite only a few students with improved grades. They indicated increased student motivation and communication, particularly from shy students and those not involved in many school activities. They also pointed to positive benefits, as one noted, from “[having a caring adult at school interested in their well-being.](#)” According to another, “[Knowing they have a go-to person has been really good for them.](#)”

Most believed the program reinforced the college-is-possible message, and that about three-fourths of the participating students want to pursue some type of post-high school education, including trade school and the military. However, they noted that poverty and family concerns are also barriers for many of these students.

Among lessons learned, one critical takeaway is the importance of mentor-mentee fit. Random assignment can pose a challenge on this issue, as mentors did not select students for the program during the RCT year. As a result, some believed the “wrong kids” were chosen, and that the program might have been more effective with other, more needy students.

Curriculum buy-in varied from mentor to mentor, based on the interviews. Their adherence to the curriculum often depended on whether the program met regularly for at least 30 minutes a week. If they had fewer regular meetings, mentors indicated it was a challenge to deliver specific lessons. Mentors also indicated that it was important to be flexible in their approach to the curriculum, as they sometimes changed a lesson plan on the spur of the moment to respond to student needs, questions, and day-to-day concerns.

Finally, although Oak Hill mentors appeared to have held the fewest meetings during the past two years, based on the interviews, exposure data described in the previous section of this report indicates that this site had the most hours of student participation during the RCT period. This finding may be due to more vigilant reporting of all student interactions, because Oak Hill mentors did report using informal contact to reach their students. Given that mentors at all

schools also had favorable views of the one-to-one check-ins with students, these findings may warrant further exploration to determine the appropriate mix of classroom presentation and individual/informal meetings in the future.

2. The Program Impact Analysis

2.1 Attitudes toward college-going and grit

The results of regression models examining differences in CGSE, CGOE, and grit scores between students in the treatment and control groups revealed no impact of the program. Table 8 summarizes the estimates for each outcome (see Appendix B for technical details and tables). The standardized effect size column illustrates the small, but not statistically significant, difference in post-intervention scores for the three outcomes. These effect sizes were much smaller than .25 (absolute value), considered by WWC as a “substantively important.”

As mentioned earlier in this report, each program impact was estimated using the coefficient derived for the treatment variable, and the effects of other important predictors (e.g., pre-intervention CGSE and CGOE, socioeconomic status) were removed.

Table 8. Program Impact for CGSE, CGOE, and Grit

Outcome Variables	Estimated Program impact	Statistical Test (alpha = 0.05)	Standardized Effect Size
CGSE	-0.06	Not Significant	-0.06
CGOE	-0.16	Not Significant	-0.15
Grit	0.00	Not Significant	0.00

Note: The grit scale was not included in the pre-test, only the post-test. As such, the analyses of grit is considered exploratory due to the fact that ICF only had access to post-intervention data.

3. Behavioral Engagement

Table 9 shows the results of regression analyses that examined two behavioral engagement outcomes—unexcused absence rates and total discipline referrals. The impact estimates were not statistically significant. The effect sizes reported indicate that from pre-intervention to post-intervention, the rate of unexcused absences increased for those in the treatment group, and the number of discipline referrals decreased, but the size of the estimates were negligibly small. Appendix B provides a table with more information on the regression results.

Table 9. Program Impact for Attendance Rate and Discipline Referrals

Outcome Variables	Estimated Program Impact	Statistical Test (alpha = 0.05)	Standardized Effect Size
Unexcused absence rate	0.00	Not Sig.	0.05
Number of discipline referrals	-0.37	Not Sig.	-0.19

4. Academic Achievement

Table 10 shows the results of regression models that examined program impacts upon cumulative GPA as measured at the end of ninth grade. We found the program impact was negative, but extremely small (-0.03) and not statistically significant. See Appendix B for more information about the results of regression models.

Table 10. Program Impact for GPA

Outcome Variables	Estimated Program Impact	Statistical Test (alpha =0.05)	Standardized Effect Size
Cumulative GPA	0.12	Not Sig.	0.11

5. Further Exploratory Analysis

In addition to the regression analyses used to measure differences between treatment and control groups, the ICF evaluation conducted exploratory analyses to identify any other results or patterns that could indicate another impact of the mentoring program or a relationship between the program and student outcomes. Results of these analyses are discussed in this section.

5.1 Subgroup impact analysis

For all six outcomes, the ICF evaluation team did not discover any statistically significant differences in outcomes between treatment and control groups. To explore potential impact among subgroups, the evaluation team conducted a subgroup impact analysis by separating the analysis sample into subsamples based on gender (male and female student samples), math achievement scores (above average and below average score student samples), and schools (three schools). No statistically significant and consistent results were observed.

5.2 Correlational analysis

As mentioned earlier, the amount of program exposure was insufficient (only 14% of students were exposed to 10 hours or more), which may explain the lack of statistically significant differences between the treatment and control groups. We conducted an additional set of exploratory analyses to see whether the amount of program exposure was correlated with student outcomes. Since treatment students were not randomly assigned to different levels of program dosage by design, the analysis is only exploratory and correlational. Because the sample size per subgroup (reported below) is relatively small, we do not provide statistical test results.

To conduct this analysis, ICF evaluators first divided the analysis sample into six subgroups of students based on group membership and hours of total exposure to mentoring. The control group consisted of students who were not exposed to the program. Groups 1 to 5 consisted of treatment students (n=90) who spent varying amounts of time participating in the mentoring program. Group 1 (zero hours) consisted of those students who chose not to use the mentoring services during the study duration (i.e., those students received zero hours of program exposure). Groups 2 to 5 consisted of students who engaged in the mentoring services from 0.1

to 4.9 hours, 5 to 9.9 hours, 10 hours to 14.9 hours, and 15 hours and above, respectively. The following list defines these subgroups:

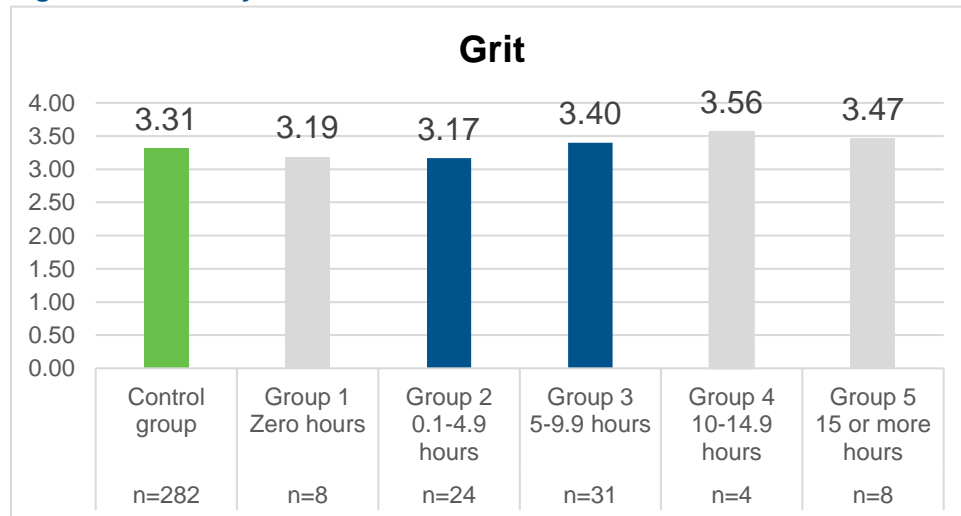
- Control group
- Group 1. Treatment students with zero hours of participation
- Group 2. Students with 0.1 to 4.9 hours of participation
- Group 3. Students with 5 to 9.9 hours of participation
- Group 4. Students with 10 to 14.9 hours of participation
- Group 5. Students with 15 or more hours of participation

Separately for each student outcome, the average scores per subgroup were calculated and compared graphically in bar charts (Figures 5 through 9). The point of analytical interest is whether the pattern suggests a correlation, either positive or negative, between program participation hours and student outcome averages. The colors of bars in the graph are green for the control group, blue for the group whose sample size is equal to or greater than 10, and gray for the group whose sample size is less than 10. Results based on the sample size of smaller than 10 students should be interpreted with caution as the reported group averages are likely unreliable.

Upon inspection of results, a relatively systematic and consistent pattern emerged for the grit scale, unexcused absence rates, total number of discipline referrals, and cumulative GPA. Specifically, the pattern suggests that greater program exposure positively correlates with favorable outcomes. The results for two college-going outcomes (CGSE and CGOE) did not show the positive and monotonic increase pattern (see Appendix B for the analysis summary of all six outcomes).

Figure 5 shows that the average grit scores increased between Group 2 and Group 3 (3.17 to 3.40; difference=0.23; standard deviation of the grit scores was 0.53). As mentioned, the gray bars should be interpreted with caution because the number of students in those groups was less than 10, and thus the subgroup averages were unreliable. Yet, there seems to be a trend of moderate or positive correlation between program exposure hours and scores on the grit scale.

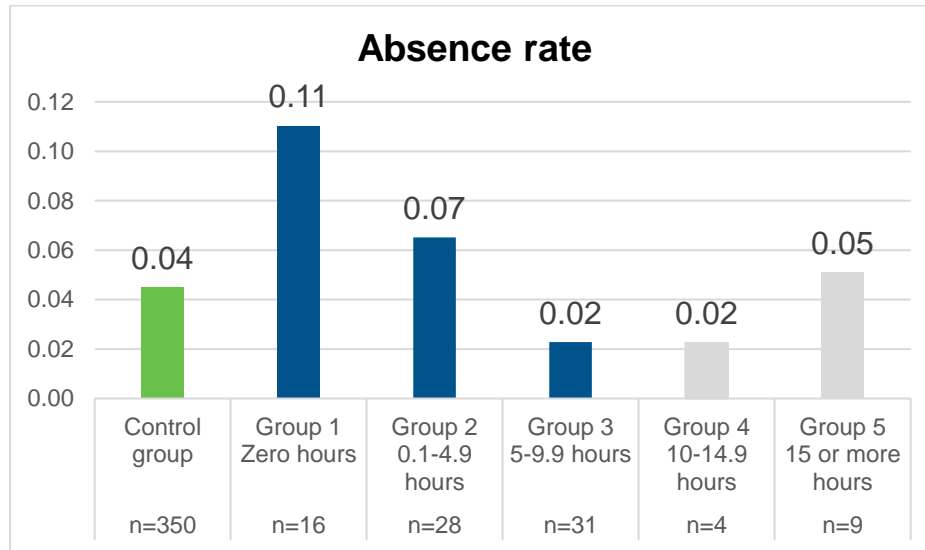
Figure 5. Grit Analysis



Note: The mean score on the grit scale was 3.31 and the standard deviation was 0.53. N=357.

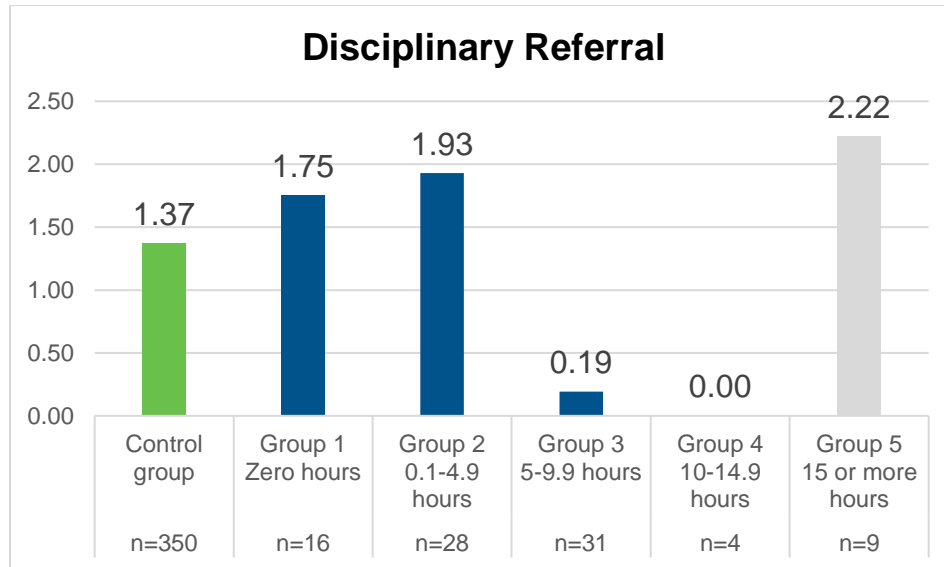
Analysis of differences in the number of discipline referrals and unexcused absence rates show similar and more consistent patterns of positive correlation between program exposure and more favorable outcome values. The absence rate is higher for Group 1 (11%) than either Group 2 (7%) or Group 3 (2%), indicating a trend that the more program hours, the lower the absence rate. The control group had a lower absence rate than Group 1, even though both groups received zero hours of program intervention. This may mean that Group 1 is a self-selected group of students who chose not to participate in the program despite the opportunity, and their orientation and attitudes may differ qualitatively from those of control group students.

Figure 6. Absence Rate Analysis



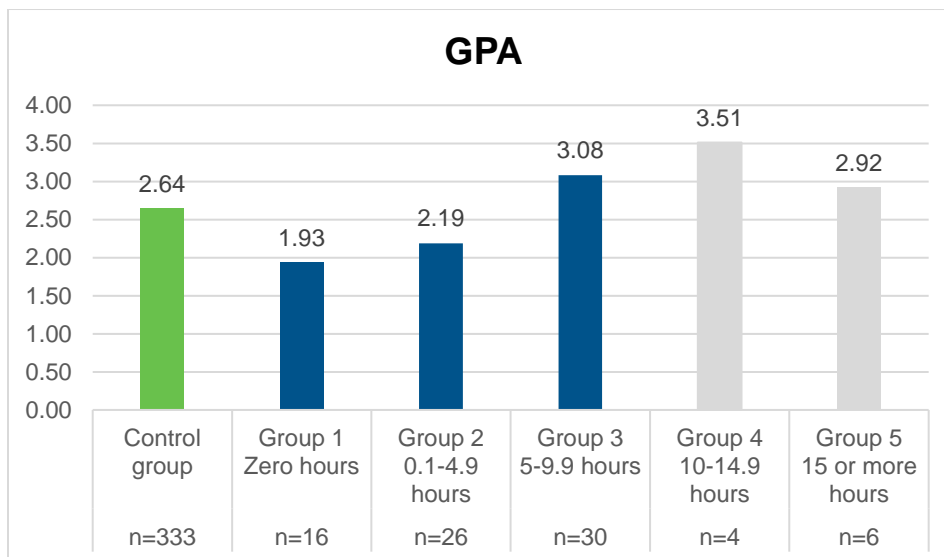
Note: The mean absence rate was 0.05 and the standard deviation was 0.07. N=438.

Likewise, we found a somewhat similar trend for disciplinary referrals. For Groups 1 and 2, the average number of disciplinary referrals was 1.75 and 1.93, respectively. While there was a slight increase between Groups 1 and 2, Group 3, which received more hours of program exposure, had an average of 0.19 discipline referrals.

Figure 7. Disciplinary Referral Analysis

Note: The mean disciplinary referral count was 1.34 and the standard deviation was 2.95. N=438.

Finally, GPA outcomes also showed the same pattern—that more exposure to mentoring seemed to be associated with better GPA outcomes. The average GPA of Group 5 (2.92) contradicts the trend, yet its sample size was smaller than Groups 1, 2, and 3, which calls for cautious interpretation.

Figure 8. GPA Analysis

Note: The mean GPA was 2.63 and the standard deviation was 1.14. N=415.

The results of these exploratory analyses demonstrated that the program exposure time was correlated positively with student outcomes. This is not a causal analysis and thus will not support the causal relationship between the program and student outcomes. The fact that the

correlation is not negative, however, suggests that the GEAR-UP program may benefit students if a sufficient amount of time is provided to program participants.

V. Discussion

As discussed earlier, WV GEAR UP regional program directors determined that at least 10 hours of mentoring constituted a minimum effective dose of the program, although we found that only 14% of treatment students met this benchmark. This finding is incredibly important as it casts doubt on whether or not the study results accurately estimate the impact of the SSS program.

Findings from the mentor interviews suggest some explanations for the low exposure to the program. The degree of exposure a student might have varied across schools and among individuals as a result of various factors. Schools varied from being unable to secure enough mentors to engaging individuals with substantial commitment to and interest in mentoring. Three mentors believed that the SSS mentoring curriculum could seem too much like a class, which negatively influenced the engagement and participation of students. For example, at one site, many students in the SSS program were also members of another program that met at the same time, and often chose the other program because it seemed less like a traditional school-day class.

Given the low level of program exposure, it is not surprising that for all six outcomes we did not find significant differences between students in the treatment group who received SSS mentoring services and those in the control group who did not.

The results of the exploratory analyses were promising and suggested that program exposure time positively correlates with favorable student outcomes related to grit, absences, discipline, and GPA. Ultimately, additional examination will be needed to understand the definitive impact of the SSS mentoring program on high-need West Virginia schools.

Findings from this study do, however, provide meaningful takeaways regarding the study design and the process of conducting an RCT on mentoring in a school setting. A fundamental question related to the study design is whether randomization was an effective approach to assess the impact of mentoring on students. For example, the literature demonstrates that the match between mentor and mentee itself can be an important predictor of success (Jekielek, Moore, & Hair, 2002). In addition, several mentors across the RCT schools believed that the selection of students through randomization resulted in the participation of the “wrong” students (i.e., students who were on the path towards successful transition to postsecondary education in any case, and thus less in need of mentoring than other students). It is important to learn more about any unintended consequences that randomization may have had on the program impact.

In addition, for a variety of logistical reasons, all this study’s exposure data were collected after the fact. This approach represents a missed opportunity to address the low dosage issue as it was happening. One recommendation would be to consider ways that future studies could collect more “just-in-time” data on program participation. With real-time data, we could monitor which students were getting adequate program hours and which were not, opening the door to addressing the issue of low exposure while it was occurring.

Despite the challenges, we note important successes related to the RCT. We were successful in completing the randomization, controlling costs through utilization of existing data, and reducing participant burden. Another major success, considering the school setting, is that we found low attrition rates for all six outcomes. A final important lesson learned is the value of monitoring a program's implementation very closely, in real time, in order to see more meaningful results.

VI. Recommendations

The evaluation team provides the following recommendations for the Commission to consider based on the findings summarized in this report.

- **Consider alternate evaluation strategies for measuring impact of the SSS mentoring program on students.** Since the average number of hours of SSS mentoring exposure for students in the treatment group was approximately half of what the WV GEAR UP regional program directors defined as an effective dose of the program (i.e., 5.5 hours out of 10 hours), the Commission may wish to consider evaluation strategies that examine impact based on those students who actually received an adequate dose of mentoring. For example, one possibility would be to conduct additional analyses looking at “treatment on the treated” or a quasi-experimental design (QED) study in which students who received an adequate dose of the SSS mentoring program are compared to a matched comparison group of students who did not receive mentoring. In addition, the Commission may also consider replicating the RCT, but including an option to provide real-time data on program participation so that the evaluation team has the opportunity to work with the Commission to address the issue of low mentoring dosage before it is too late.
- **Explore grit scale reliability further and consider alternate ways of measuring grit.** The eight-item grit questionnaire included in the student survey had low reliability (.50) based on the year 4 student responses for those students participating in the RCT. Although the low level of SSS program exposure is likely the primary reason why impact was not observed between treatment and control groups, it does not help to have unreliable instruments for measuring impact. One possible reason for the low reliability is that half of the items on the grit scale required reverse coding, as described in the Data Sources section. While the reverse coding is likely intended to increase the internal reliability of the scale, it may have been confusing for students, particularly students with lower levels of literacy. In addition, as the grit scale was included at the end of the survey, students who were eager to finish the survey may have moved quickly through the items and not read the wording closely. There are other measures of grit, including a 12-item grit scale; it may be worth exploring whether such alternate measures have more reliable results (Duckworth et al., 2007). Alternatively, it is possible that moving the placement of the grit scale in the survey could also lead to greater reliability. Finally, it may be worth exploring the reliability of the grit scale and the other college-going scales—CGSE and CGOE—among all year 4 student survey respondents, not just those in the RCT. Including a greater number of respondents in this analysis will provide a more accurate depiction of the reliability of the scales for the given audience (i.e., grade 10 students in the WV GEAR UP high schools).
- **Consider a time frame longer than one academic year for future studies on mentoring.** Not only would additional time allow student/mentor matches to build relationships over a

longer span of time, but it would also allow for greater program exposure. The exploratory analyses suggested that the greater the program exposure, the more favorable outcomes. Literature also suggests that the duration of mentoring may affect the impacts of a mentoring program on students (Tierney & Grossman, 2000; Herrera, 2004). Extending the time frame of the study would increase the likelihood of detecting program impacts.

- **Explore revising the SSS mentoring curriculum to be more engaging to students.** As noted in the mentor interview findings, three of the seven mentors felt that the SSS mentoring sessions could seem too much like a class, with students becoming less engaged and even attending fewer mentoring sessions as a result (as in the case of the one school that held a competing GEAR UP activity at the same time as the SSS program). Mentors had positive feedback, however, about the SSS group activities, which they believed distinguished SSS from a typical class and encouraged team building and volunteerism. Potential revisions to the SSS curriculum to better engage students, such as placing a greater emphasis on the SSS group activities, for example, could lead to increased student participation in the SSS program. Considering the results of the exploratory correlational analysis, which suggested that greater exposure to the mentoring program may correlate with favorable student outcomes related to grit, absences, discipline, and GPA, increased student participation in mentoring may be beneficial to students with respect to these outcomes. Ultimately, interview findings from all mentors, not just those from the RCT sites—which will be included in the forthcoming Year 4 Annual Evaluation Report—should be reviewed before considering any potential curriculum revisions. On a related note, future studies on mentoring may wish to look at the format/content of mentoring—whether the mentoring program is more structured (like a class) versus more flexible and how that may affect mentoring impacts.
- **Consider providing guidance to schools on how to carve out time for SSS mentoring sessions.** As noted in the interviews with mentors, two of the three sites encountered challenges carving out a regular time for students to meet with mentors during the 2016-2017 year. Mentors who were teachers, however, were more likely to indicate that they had sufficient time to meet regularly with students, particularly if their school set aside dedicated time for the SSS program, such as an advisory or lunch period. The Commission may consider providing guidance to schools about how to set aside dedicated time for mentoring.
- **Share lessons learned from the process of conducting an RCT in the three West Virginia schools.** Various elements of the SSS evaluation design were successful and could be replicated in other contexts, including, randomization, controlling costs through utilization of existing data, and reducing participant burden. Some elements could be improved; for example, instead of collecting program exposure data at the end of the study, collecting these data on an ongoing basis would allow the evaluation team to address low rates of exposure in time to address the problem. This would help to increase the likelihood of being able to measure program impact through the study design. Ultimately, these are valuable lessons that would be beneficial to share with other education program leaders and evaluators—particularly as RCTs offer the most rigorous evaluation of program impact and offer causal evidence. Potential forums for sharing these lessons learned include conferences (e.g., the National Council for Community and Education Partnerships Annual Conference, the American Evaluation Association’s Annual Meeting, the American

Educational Research Association's Annual Meeting) and a range of education research and evaluation journals.

VII. References

- Bruce, M., & Bridgeland, J. (2014). *The mentoring effect: Young people's perspectives on the outcomes and availability of mentoring*. National Mentoring Partnership and Civic Enterprises. Retrieved from: http://www.mentoring.org/images/uploads/Report_TheMentoringEffect.pdf
- Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (Grit-S). *Journal of Personality Assessment*, 91, 166-174. https://globaled.gse.harvard.edu/files/geii/files/validation_grit_scale_duckworth_jpa_m.fiqueroa-2.pdf
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 9, 1087-1101. https://www.researchgate.net/publication/6290064_Grit_Perseverance_and_Passion_for_Long-Term_Goals
- Gibbons, Melinda M. (2005). *College-going beliefs of prospective first-generation college students: Perceived barriers, social supports, self-efficacy, and outcome expectations*. Directed by L. DiAnne Borders. The University of North Carolina at Greensboro. Retrieved from <http://libres.uncg.edu/ir/uncg/f/umi-uncg-1049.pdf>
- Herrera, C., DuBois, D., & Grossman, J. (2013). *The role of risk: Mentoring experiences and outcomes for youth with varying risk profiles*. MDRC. Retrieved from: https://www.mdrc.org/sites/default/files/Role%20of%20Risk_Final-web%20PDF.pdf
- Herrera, C. (2004). *School-based mentoring: A closer look*. Philadelphia: Public/Private Ventures. Retrieved from http://dropoutprevention.org/wp-content/uploads/2015/05/PPV_School_Based_Mentoring_A_Closer_Look_2004.pdf
- Jekielek, S., Moore, K., & Hair, E. (2002). *Mentoring programs and youth development: A synthesis*. Child Trends and Edna McConnell Clark Foundation. Retrieved from: <https://www.childtrends.org/publications/mentoring-programs-and-youth-development-a-synthesis/>
- National Center for Education Statistics, *The condition of education 2017*. Washington, DC: National Center for Education Statistics, U.S. Department of Education. Retrieved from: <https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2017144>
- Tierney, J.P., and Grossman, J.B. (2000). *Making a difference: An impact study of Big Brothers Big Sisters*. Philadelphia: Public/Private Ventures. Retrieved from <https://www.issuelab.org/resource/making-a-difference-an-impact-study-of-big-brothers-big-sisters-re-issue-of-1995-study.html>
- United States Department of Education, Institute of Education Sciences. (2013), What Works Clearinghouse Evidence Standards, Version 3.0. Retrieved from <http://whatworks.ed.gov>
- West Virginia Department of Education, (2018). *ZoomWV data dashboard*. Retrieved from: <https://zoomwv.k12.wv.us/Dashboard/portalHome.jsp>.

West Virginia Higher Education Policy Commission and West Virginia Community and Technical College System, (2018). *West Virginia higher education report card*. Retrieved from: http://www.wvhepc.edu/wp-content/uploads/2018/01/ReportCard2017_Final_10April2018.pdf.

Appendix A: Opt-Out/Permission Forms and Instrument

Appendix A contains the following documents:

- Student Success Society Parent/Guardian Opt-Out Form
- Year 4 Student Survey Parent/Guardian Permission Form
- Year 4 Student Survey Instrument

Appendix B: Technical Detail

1. Attrition Analysis

From left to right, Table B1 summarizes information for the whole sample, the treatment group, and the control group. Within each of these segments, the table reports from left to right a) original sample size, b) analysis sample size, and c) attrition rate. Rows represent six student outcomes examined in this report. Attrition rates were calculated for the whole sample, the treatment group sample, and the control group sample. To decide whether a study sample suffered a high level of attrition, WWC takes into consideration the overall attrition rate and differential attrition rates between treatment and control groups. We rated the attrition for all six outcomes as low, and thus we can have confidence in the integrity of the RCT design and its ability to estimate the program impact without bias.

Table B1. Overall and Differential Attrition Rates of Study Samples per Outcome

Outcomes	The Whole Sample			Treatment Group			Control Group			Differential Attrition	WWC Attrition Level
	Original Sample	Analysis Sample	Attrition Rate	Original Sample	Analysis Sample	Attrition Rate	Original Sample	Analysis Sample	Attrition Rate		
CGSE	450	340	24%	90	72	20%	360	268	26%	6%	Low
CGOE	450	337	25%	90	71	21%	360	266	26%	5%	Low
Grit	450	338	25%	90	72	20%	360	266	26%	6%	Low
Unexcused absence rate	450	423	6%	90	86	4%	360	337	6%	2%	Low
Disciplinary referral	450	423	6%	90	86	4%	360	337	6%	2%	Low
GPA	450	400	11%	90	80	11%	360	320	11%	0%	Low

Note: WWC's liberal definition of attrition was used to determine whether the data had a low or high level of attrition.

2. Baseline Equivalence Analysis

When an RCT study suffers high attrition, WWC requires that a study should establish baseline equivalence using prior-to-the-intervention statistics. This study had low attrition, and thus there was no need to establish that treatment and control groups were equivalent at baseline. For reference purposes, however, Table B2 summarizes the pretest characteristics of the two groups in relation to the most relevant prior-to-the-intervention variables and pretest math standardized test scores. The group difference of all variables considered was sufficiently small such that none of the models would have been rated as failing to establish baseline equivalence. When standardized group difference of a variable (e.g., pretest math score) was greater than 0.05 and smaller than .25, that variable must be included in the statistical model for covariate adjustment. This study was a low attrition RCT and thus, under WWC Evidence Standards, covariate adjustment was not necessary. Yet the proposed statistical model did include these variables as they were important predictors of the outcomes.

Table B2. Baseline Equivalence Analysis for Six Outcomes

	Treatment Group			Control Group				
	N.	Mean	SD	N.	Mean	SD	Standardized Difference (Hedge's g)	WWC Adjustment
CGSE Analysis								
Pretest CGSE	72	3.56	0.96	268	3.43	0.97	0.13	Covariate adjustment necessary
Pretest math test scores	72	2518.38	110.2	268	2516.59	99.77	0.02	
CGOE Analysis								
Pretest CGOE	71	3.71	1.06	266	3.55	1.02	0.15	Covariate adjustment necessary
Pretest math test scores	71	2518.24	110.98	266	2517.47	99.5	0.01	
Grit Analysis								
Pretest grit scores	Not available							N/A
Pretest CGSE	72	3.56	0.96	266	3.44	0.98	0.12	Covariate adjustment necessary
Pretest math test scores	72	2518.38	119.2	266	2517.35	99.73	0.01	
Discipline Referral Analysis								
Pretest discipline referral	86	1.95	4.63	337	1.68	3.72	0.07	Yes
Pretest math test scores	86	2505.2	114.09	337	2506.34	99.49	-0.01	
Unexcused Absence Analysis								
Pretest absence rate	86	0.04	0.04	337	0.04	0.04	0.07	Covariate adjustment necessary
Pretest math test scores	86	2505.2	114.09	337	2506.34	99.49	-0.01	
GPA Analysis								
Pretest GPA	Not available							N/A
Pretest math test scores	80	2488.85	99.52	320	2499.77	95.04	-0.11	Covariate adjustment necessary

Note: "Covariate adjustment necessary" in the WWC adjustment column means that a variable had a greater than 0.05 and smaller than 0.25 standardized difference (absolute value of Hedge's g statistic) and thus a high attrition RCT would have to include the variable as a predictor to adjust for the difference. When used in the statistical models, pretest math scores were centered around the mean (2506.11; this value was subtracted from each of the scores).

3. Multivariate Regression Analysis

The equation below represents general modeling strategies with the dummy variable indicating the treatment group (vs. the control group).

$$Posttest_i = \beta_0 + \beta_1 * Treatment + \beta_2 * Pretest + \beta_3 * Covariate + \dots + r_i$$

The program impact will be estimated by β_1 , the impact coefficient, which was adjusted for all the variables included in the model. The covariates were pretest mathematics standardized test scores (proxy of student academic ability), race and ethnicity variables (White, Black, and other race; White is the omitted category), low socioeconomic status (SES) (recipient of Supplemental Nutrition Assistance Program vs. non-recipient), special education status, and schools (Mingo Central High School, Nicholas County High School, and Oak Hill High School; Oak Hill High School is the omitted category). The intercept value (β_0) corresponds to the expected outcome value of a person who had zeros on all predictors (e.g., control students, White students). Because math test scores were large values and a score of zero did not exist, the scores were centered around the sample mean (2506.11). The intercept, therefore, is an expected outcome value for a student who had the average pretest math score.

Table B3. Summarizes All Covariates Used in the Statistical Models.

	Covariates	Notes
CGSE	Pretest-CGSE, pretest math score, gender, race (Black, White, other race), student status (low SES, special education), special education, school (n=3).	
CGOE	Pretest-CGOE, pretest math score, gender, race (Black, White, other race), student status (low SES, special education), special education, school (n=3).	
Grit	Pretest-CGSE, pretest math score, gender, race (Black, White, other race), student status (low SES, special education), special education, school (n=3).	Pretest grit score was not available.
Unexcused absence rate	Pretest-absence rate, pretest math score, gender, race (Black, White, other race), student status (low SES, special education), special education, school (n=3).	
Disciplinary referral	Pretest-disciplinary referral, pretest math score, gender, race (Black, White, other race), student status (low SES, special education), special education, school (n=3).	
GPA	Pretest math score, gender, race (Black, White, other race), student status (low SES, special education), special education, school (n=3).	Pretest GPA was not available.

Table B4. Results of Multivariate Regression Models for CGSE and CGOE Outcomes

	CGSE					CGOE				
	Coefficient	Error	P.	sig	Standardized Coefficient	Coefficient	Error	P.	sig	Standardized Coefficient
Intercept	1.54	0.19	0.00	***	-1.97	1.64	0.21	0.00	***	-1.76
Treatment	-0.06	0.10	0.53		-0.06	-0.16	0.11	0.14		-0.15
Pretest CGSE or CGOE	0.59	0.04	0.00	***	0.61	0.57	0.05	0.00	***	0.54
Pretest math score	0.00	0.00	0.01	**	0.00	0.00	0.00	0.05	*	0.00
Male	-0.16	0.08	0.04	*	-0.17	-0.11	0.09	0.24		-0.10
Black	0.32	0.18	0.08		0.33	0.19	0.21	0.37		0.17
Other	-0.14	0.37	0.71		-0.14	-0.21	0.42	0.61		-0.20
Low SES	-0.17	0.08	0.03	*	-0.18	-0.22	0.09	0.02	*	-0.20
Special education	-0.05	0.14	0.73		-0.05	-0.15	0.17	0.38		-0.14
Mingo Central High School	-0.07	0.10	0.46		-0.08	-0.11	0.12	0.33		-0.11
Nicholas County High School	0.15	0.10	0.12		0.15	0.14	0.11	0.21		0.13
Outcome variance	0.52					0.68				
Original variance	0.96					1.14				
Variance explained	0.46					0.40				

Note: The sample size is 340 for CGSE and 337 for CGOE. The reference categories are White students and Oak Hill High School students. Asterisks: *** if $p < 0.001$, ** if $p < 0.01$, * if $p < 0.05$.

Table B5. Results of Multivariate Regression Models for Grit Scale

Effect	Coefficient	Error	P.	sig	Standardized Coefficient
Intercept	2.90	0.13	0.00	***	-0.79
Treatment	0.00	0.07	0.99		0.00
Pretest CGSE	0.12	0.03	0.00	***	0.23
Pretest math score	0.00	0.00	0.11		0.00
Male	-0.01	0.06	0.90		-0.01
Black	0.01	0.13	0.92		0.02
Other	0.12	0.26	0.63		0.23
Low SES	-0.03	0.06	0.57		-0.06
Special education	-0.27	0.10	0.01	**	-0.51
Mingo Central High School	-0.01	0.07	0.85		-0.02
Nicholas County High School	0.09	0.07	0.19		0.17
Outcome variance	0.26	0.02			0.91
Original variance	0.28				
Variance explained	0.09				

Notes: The sample size is 338. The reference categories are White students and Oak Hill High School students. Asterisks: *** if $p < 0.001$, ** if $p < 0.01$, * if $p < 0.05$.

Table B6. Results of Multivariate Regression Models for Unexcused Absence Rate and Discipline Referral

	Coeffi- cient	Error	P.	sig	Standar- dized Coeffi- cient	Coeffi- cient	Error	P.	sig	Standar- dized Coeffi- cient
Intercept	0.00	0.01	0.49		-0.59	0.93	0.27	0.00	***	-0.11
Treatment	0.00	0.01	0.63		0.05	-0.37	0.28	0.19		-0.14
Pretest absence rate or disciplinary referral	0.81	0.07	0.00	***	11.90	0.19	0.03	0.00	***	0.07
Pretest math score	0.00	0.00	0.00	**	0.00	-0.01	0.00	0.00	***	-0.00
Male	0.00	0.01	0.54		-0.05	0.27	0.23	0.24		0.10
Black	0.01	0.01	0.65		0.08	1.02	0.53	0.06		0.39
Other	-0.02	0.03	0.39		-0.35	-1.65	1.16	0.16		-0.64
Low SES	0.01	0.01	0.10		0.13	0.27	0.23	0.24		0.10
Special education	0.00	0.01	0.72		0.05	-0.25	0.40	0.53		-0.10
Mingo Central High School	0.02	0.01	0.01	*	0.24	0.24	0.29	0.40		0.09
Nicholas County High School	0.00	0.01	0.47		-0.07	-1.01	0.28	0.00	***	-0.39
Outcome variance	0.00	0.00				5.25				
Original variance	0.00					0.46				
Variance explained	0.36					0.22				

Notes: The sample size is 423 for both absence rate and disciplinary referral. The reference categories are White students and Oak Hill high school students. Asterisks: *** if $p < 0.001$, ** if $p < 0.01$, * if $p < 0.05$.

Table B7. Results of Multivariate Regression Models for GPA

	Coefficient	Error	P.	sig	Standardized Coefficient
Intercept	3.35	0.11	0.00	***	0.61
Treatment	0.12	0.11	0.29		0.11
Pretest math score	0.01	0.00	0.00	***	0.01
Male	-0.44	0.09	0.00	***	-0.40
Black	-0.26	0.22	0.24		-0.24
Other	0.30	0.44	0.49		0.28
Low SES	-0.42	0.09	0.00	***	-0.38
Special education	0.12	0.15	0.43		0.11
Mingo Central High School	-0.54	0.12	0.00	***	-0.49
Nicholas County High School	-0.02	0.11	0.85		-0.02
Outcome variance	0.77	0.05			0.63
Original variance	1.21				
Variance explained	0.37				

Notes: The sample size is 400. The reference categories are White students and Oak Hill High School students. Asterisks: *** if $p < 0.001$, ** if $p < 0.01$, * if $p < 0.05$.

4. Descriptive Statistics of Analysis Samples

Table B8. Descriptive Statistics for the CGSE Impact Analysis

	N	Min	Max	Mean	SD
CGSE	340	1	5	3.46	0.98
Treatment	340	0	1	0.21	0.41
Pretest CGSE	340	1	5	3.46	0.97
Pretest math score	340	-271.11	317.89	10.86	101.9
Male	340	0	1	0.51	0.50
Black	340	0	1	0.05	0.22
Other (Race)	340	0	1	0.01	0.11
Low SES	340	0	1	0.51	0.50
White	340	0	1	0.94	0.25
Special education	340	0	1	0.09	0.29
Mingo Central High School	340	0	1	0.31	0.46
Nicholas County High School	340	0	1	0.36	0.48
Oak Hill High School	340	0	1	0.33	0.47

Note: The reference categories were White students and Oak Hill High School students.

Table B9. Descriptive statistics for the CGOE analysis

	N	Min	Max	Mean	SD
CGOE	337	1	5	3.53	1.07
Treatment	337	0	1	0.21	0.41
Pretest CGSE	337	1	5	3.59	1.03
Pretest math score	337	-271.11	317.89	11.52	101.86
Male	337	0	1	0.5	0.5
Black	337	0	1	0.05	0.23
Other (Race)	337	0	1	0.01	0.11
Low SES	337	0	1	0.51	0.50
White	337	0	1	0.93	0.25
Special education	337	0	1	0.09	0.29
Mingo Central High School	337	0	1	0.31	0.46
Nicholas County High School	337	0	1	0.37	0.48
Oak Hill High School	337	0	1	0.33	0.47

Note: The reference categories were White students and Oak Hill High School students.

Table B10. Descriptive Statistics for the Grit Scale Analysis

	N	Min	Max	Mean	SD
Grit	338	1	5	3.32	0.53
Treatment	338	0	1	0.21	0.41
Pretest CGSE	338	1	5	3.46	0.97
Pretest math score	338	-271.11	317.89	11.46	101.88
Male	338	0	1	0.51	0.5
Black	338	0	1	0.05	0.22
Other	338	0	1	0.01	0.11
Low SES	338	0	1	0.51	0.50
White	338	0	1	0.93	0.25
Special education	338	0	1	0.09	0.29
Mingo Central High School	338	0	1	0.31	0.46
Nicholas County High School	338	0	1	0.36	0.48
Oak Hill High School	338	0	1	0.33	0.47

Note: The reference categories were White students and Oak Hill High School students.

Table B11. Descriptive Statistics for the Unexcused Absence Rate Analysis

	N	Min	Max	Mean	SD
Post-test absence rate	423	0	0.65	0.05	0.07
Treatment	423	0	1	0.2	0.4
Pretest absence rate	423	0	0.29	0.04	0.04
Pretest math score	423	-353.11	317.89	0	102.49
Male	423	0	1	0.52	0.50
Black	423	0	1	0.05	0.22
Other	423	0	1	0.01	0.1
Low SES	423	0	1	0.54	0.50
White	423	0	1	0.94	0.24
Special education	423	0	1	0.1	0.3
Mingo Central High School	423	0	1	0.32	0.47
Nicholas County High School	423	0	1	0.35	0.48
Oak Hill High School	423	0	1	0.33	0.47

Note: The reference categories were White students and Oak Hill High School students.

Table B12. Descriptive Statistics for Disciplinary Referral Analysis

	N	Min	Max	Mean	SD
Post-test disciplinary referral	423	0	18	1.21	2.59
Treatment	423	0	1	0.2	0.4
Pretest disciplinary referral	423	0	31	1.73	3.92
Pretest math score	423	-353.11	317.89	0.00	102.49
Male	423	0	1	0.52	0.50
Black	423	0	1	0.05	0.22
Other	423	0	1	0.01	0.10
Low SES	423	0	1	0.54	0.50
White	423	0	1	0.94	0.24
Special education	423	0	1	0.10	0.30
Mingo Central High School	423	0	1	0.32	0.47
Nicholas County High School	423	0	1	0.35	0.48
Oak Hill High School	423	0	1	0.33	0.47

Note: The reference categories were White students and Oak Hill High School students.

Table B13. Descriptive Statistics for GPA Analysis

Variable	N	Min	Max	Mean	SD
Post-test GPA	400	0	4.3784	2.69	1.1
Treatment	400	0	1	0.2	0.4
Pretest math score	400	-353.11	317.89	-8.53	95.93
Male	400	0	1	0.52	0.5
Black	400	0	1	0.04	0.2
Other	400	0	1	0.01	0.1
Low SES	400	0	1	0.54	0.5
White	400	0	1	0.95	0.22
Special education	400	0	1	0.1	0.3
Mingo Central High School	400	0	1	0.34	0.47
Nicholas County High School	400	0	1	0.37	0.48
Oak Hill High School	400	0	1	0.3	0.46

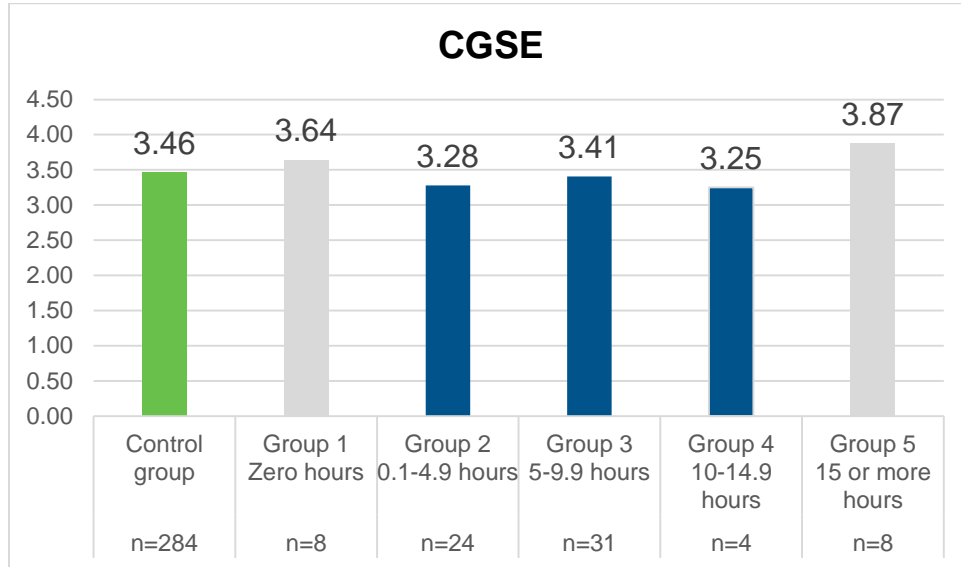
Note: The reference categories were White students and Oak Hill High School students.

5. Correlational Analysis of Program Exposure Hours and Student Outcomes

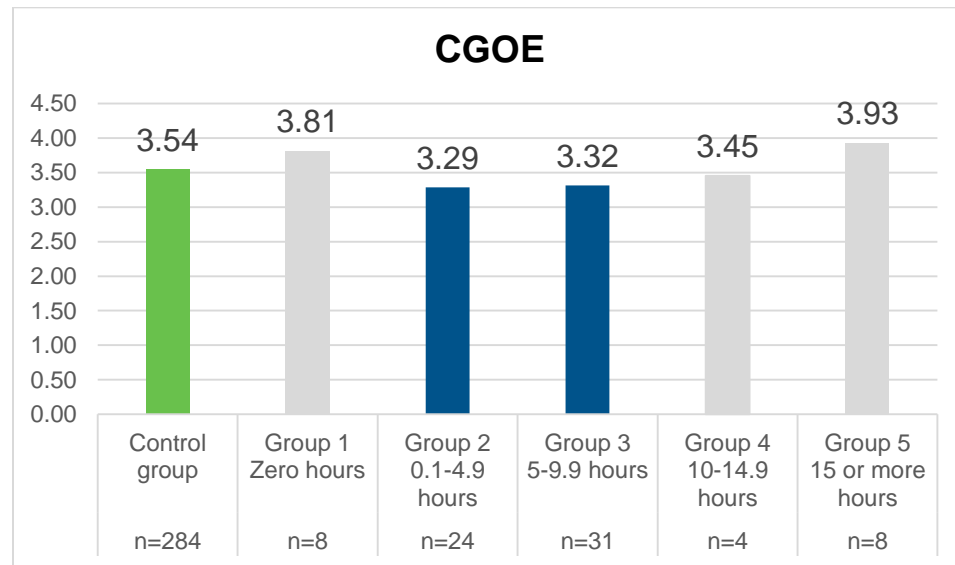
The analysis goal was to assess the degree to which program hours is associated with student outcomes. The program exposure hours were recorded in hours and minutes. As detailed in the report, the six outcomes were CGSE, CGOE, Grit, unexcused absence rate, disciplinary referral, and GPA. As shown in Table B14, students were grouped into six subgroups based on the treatment status and the amount of program exposure time. Figures B1 and B2 graphically summarize the results for CGSE and CGOE, which were not presented in the Results section because the patterns did not support the expectation that the program exposure time and student outcomes were correlated.

Table B14. Students' Program Exposure Time and Outcome Average Scores

1. Attitudes about College Going									
	CGSE				CGOE			GRIT	
Subgroups	N	Mean	SD	N	Mean	SD	N	Mean	SD
Control group	284	3.46	0.95	284	3.54	1.02	282	3.31	0.54
Group 1: Zero hours	8	3.64	0.89	8	3.81	0.89	8	3.19	0.25
Group 2: 0.1-4.9 hours	24	3.28	1.13	24	3.29	1.20	24	3.17	0.45
Group 3: 5-9.9 hours	31	3.41	1.00	31	3.32	1.27	31	3.40	0.52
Group 4: 10-14.9 hours	4	3.25	1.68	4	3.45	1.72	4	3.56	0.77
Group 5: 15 or more hours	8	3.87	1.07	8	3.93	1.08	8	3.47	0.62
2. Behavioral Engagement									
Unexcused absence rate				Disciplinary referral					
Subgroups	N	Mean	SD	N	Mean	SD			
Control group	350	0.04	0.06	350	1.37	2.95			
Group 1: Zero hours	16	0.11	0.17	16	1.75	4.46			
Group 2: 0.1-4.9 hours	28	0.07	0.10	28	1.93	3.38			
Group 3: 5-9.9 hours	31	0.02	0.02	31	0.19	0.48			
Group 4: 10-14.9 hours	4	0.02	0.03	4	0.00	0.00			
Group 5: 15 or more hours	9	0.05	0.05	9	2.22	3.07			
3. Academic Achievement									
Subgroups	N	Mean	SD						
Control group	333	2.64	1.17						
Group 1: Zero hours	16	1.93	1.17						
Group 2: 0.1-4.9 hours	26	2.19	1.02						
Group 3: 5-9.9 hours	30	3.08	0.67						
Group 4: 10-14.9 hours	4	3.51	0.61						
Group 5: 15 or more hours	6	2.92	0.47						

Figure B1. CGSE Analysis

Note: The mean CGSE was 3.46 and the standard deviation was 0.97. N=359.

Figure B2. CGOE Analysis

Note: The mean CGOE was 3.52 and the standard deviation was 1.06. N=359.