Independent Evaluation of Lone Star AP CSP: Final Report

Deborah J. Holtzman, Kathleen T. Jones, & Charles Blankenship

Report to the National Math + Science Initiative

Regarding Grant U411C190037

August 2025

American Institutes for Research® | AIR.ORG

AIR®
Advancing Evidence.
Improving Lives.

Contents

Exhibits

Executive Summary

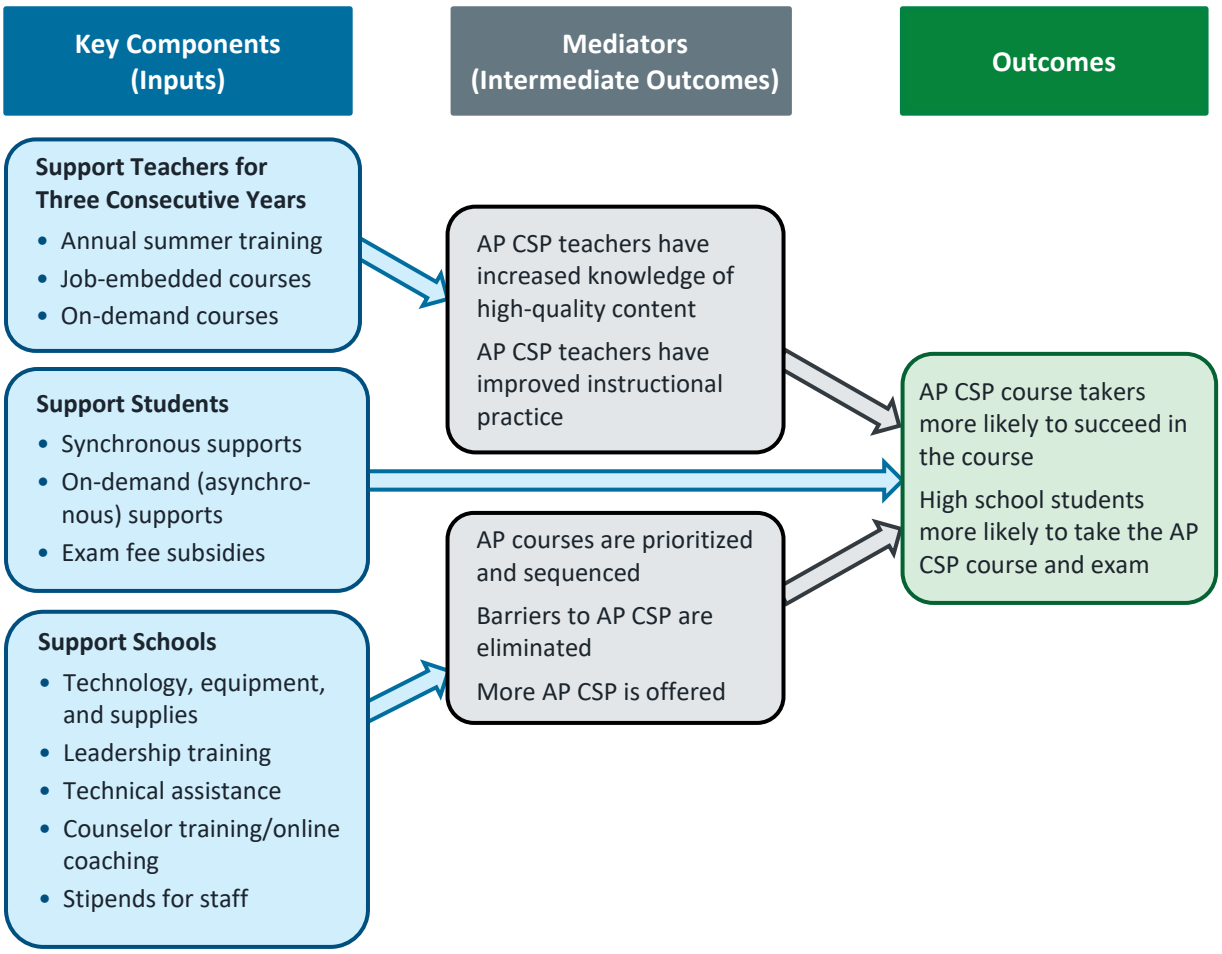
This executive summary summarizes the findings of the independent evaluation of the Lone Star Advanced Placement (AP) Computer Science Principles (CSP) program. This program, funded by an Education and Innovation Research (EIR) early-phase grant from the U.S. Department of Education, was developed by the National Math + Science Initiative (NMSI) to increase student access to and success in the AP CSP course in Texas high schools. The American Institutes for Research® (AIR®) conducted the evaluation of the program, including studies of implementation, impact on teachers' instruction, and impact on students' participation in and performance on the AP CSP exam.

Program Description and Logic Model

The Lone Star AP CSP program began in summer 2020. It was designed as a three-year program whose key components were teacher professional development and supports for students and schools. The logic model presented in Exhibit ES.1 depicts how these components were theorized to bring about the desired outcomes. As shown in the logic model, the teacher professional development component was theorized to improve teachers' content knowledge and instructional practice related to AP CSP, which in turn was theorized to improve students' success in the AP CSP course; the supports for students were theorized to directly improve students' success in the AP CSP course as well as increase the likelihood that they would take the course and the AP exam; finally, the supports for schools were designed to encourage schools to increase their AP CSP offerings and eliminate barriers to taking AP CSP, which in turn were theorized to improve students' likelihood of taking the course.

Thirty-seven schools, all in Texas (by design), were recruited by NMSI in spring 2020 to participate in the program. Over the three years of the program, a total of 61 AP CSP teachers participated, 27 of whom participated for all three years.

Exhibit ES.1. Lone Star AP CSP Logic Model



Evaluation Research Questions

AIR conducted the independent evaluation of the Lone Star AP CSP program. Using administrative and survey data, the evaluation addressed the following research questions (RQs):

1. Was the Lone Star AP CSP program implemented with fidelity in
    a. its first year (the 2020–21 school year)?
    b. its second year (the 2021–22 school year)?
    c. its third year (the 2023–23 school year)?
2. To what extent did Lone Star AP CSP program participants make use of the program and find it useful?
3. Did participation in the Lone Star AP CSP program improve AP CSP teachers' instructional practices in the AP CSP course, compared to AP CSP teachers in the business-as-usual condition, after one year, after two years, and after three years?

4. What was the impact of one year of participation in the Lone Star AP CSP program, compared to business-as-usual AP CSP participation, on high schools' rates of (a) AP CSP exam takers receiving a qualifying score on the AP CSP exam, and (b) students taking the AP CSP exam?
5. What was the impact of *two* years of participation in the Lone Star AP CSP program, compared to business-as-usual AP CSP participation, on the same outcomes?
6. What was the impact of *three* years of participation in the Lone Star AP CSP program, compared to business-as-usual AP CSP participation, on the same outcomes?

The remainder of this executive summary briefly describes the methods used to address the research questions and summarizes the findings.

Fidelity of Implementation (RQ 1)

Our approach to measuring fidelity, which involved close collaboration with NMSI, involved identifying the main elements of the program in each of its three key components and setting criteria by which to score the implementation of each element and, subsequently, of each of the three components as a whole. We measured fidelity of implementation at the program level; accordingly, the fidelity measures cover the entire sample of schools and teachers participating in Lone Star AP CSP. Most of the data used in our fidelity calculations were administrative data provided by NMSI. We assessed fidelity of implementation separately in each of the three years studied in AIR's evaluation.

The key findings from the fidelity-of-implementation analysis are presented in Exhibit ES.2 and can be summarized as follows:

- The teacher supports component was not implemented with fidelity in any of the three years.
- The student supports component was implemented with fidelity in all three years.
- The school supports component was implemented with fidelity in the first year but not in the second or third years.

Exhibit ES.2. Summary of Lone Star AP CSP Fidelity of Implementation

Component	Year 1 (2020–21)	Year 2 (2021–22)	Year 3 (2022–23)
Teacher supports	Not met	Not met	Not met
Student supports	Met	Met	Met
School supports	Met	Not met	Not met

In summary, there was no year in which all three components were implemented with fidelity, and in each of Years 2 and 3, only one of the three components was implemented with fidelity. Moreover, in none of the years was the teacher supports component—which was the program element that, as a three-year arc, made the Lone Star AP CSP program innovative and different from NMSI's other offerings—implemented with fidelity. The overall low levels of implementation fidelity are key context for interpreting the impact results presented below.

Participant Engagement and Perceptions (RQ 2)

In May of each of the three years of implementation of the Lone Star AP CSP program, NMSI invited the Lone Star AP CSP teachers to complete a survey developed and administered by AIR. Each year's survey, conducted online, asked the teachers about their participation in the program activities over the course of the preceding school year, as well as their perceptions of usefulness of each activity they participated in and barriers to participation. The surveys also asked about the participation of the teachers' AP CSP students, given that the program includes supports and resources for students along with professional development and supports for teachers.

The key findings about participant engagement and perceptions can be summarized as follows:

- Based on teachers' self-report on surveys, teacher participation in Lone Star AP CSP activities was relatively low in Year 1 and somewhat higher in Year 2. There was little change from Year 2 to Year 3 in participation levels, but perceptions of usefulness of the teacher activities increased across the three years.
- Use of the student resources, as reported by teachers, increased from year to year, as did teacher perceptions of the usefulness of these resources.
- Lack of awareness about the program's supports and resources was a commonly cited reason for nonuse of particular activities in all three years, but more so in Year 1 than Year 2, and more so in Year 2 than Year 3.
- At the close of the program in Year 3, most teachers (60%) reported that overall, they had found the program to be at least moderately useful.
- The subset of teachers who participated in all three years of the program had mixed opinions about the cohesiveness of the program across the three years, such as how well it implemented the envisioned three-year arc of professional development.

Selected additional findings from the study of participant engagement and perceptions are presented in Exhibit ES.3.

Exhibit ES.3. Selected Findings on Lone Star AP CSP Participant Engagement and Perceptions

Measure	Year 1 (2020–21)	Year 2 (2021–22)	Year 3 (2022–23)
Average number of the six teacher activities teachers reported participating in	2.7 (*n* = 28)	3.6 (*n* = 25)	3.6 (*n* = 30)
Percentage of teachers who reported that the Summer Series was at least moderately useful	50% (*n* = 26)	67% (*n* = 18)	73% (*n* = 22)
Percentage of teachers who reported any use of the student resources	36% (*n* = 28)	56% (*n* = 25)	60% (*n* = 30)
Percentage of teachers who reported that the student resources were at least moderately useful	30% (*n* = 10)	64% (*n* = 14)	72% (*n* = 18)
Percentage of teachers who reported finding the program at least moderately useful	(not asked)	(not asked)	60% (*n* = 30)

A serious caveat of these findings, however, is that they are based on teachers' self-report on surveys, and the self-report may have been inaccurate. Moreover, not all Lone Star teachers responded to the surveys, and those who did may not have been representative of the full sample. Thus, these findings were intended mainly as formative feedback for NMSI.

Impact on Instructional Practices (RQ 3)

As shown in the logic model, instructional practice in the AP CSP course was one of the key mediators by which the program aimed to have an impact on students. Accordingly, AIR's evaluation studied the impact of the program on teachers' practice. We examined three instructional practice outcomes that were derived from a study-administered survey of AP CSP teachers (including Lone Star AP CSP teachers as well as a volunteer sample of AP CSP teachers not in the program):

- Engaging Students in Practices of Computer Science
- Encouragement of Cognitive Problem-Solving Strategies
- Use of Pedagogically Based Learning Strategies

Separately in each of the three outcome years, the instructional practices impact study used a quasi-experimental matched comparison group design and an analysis-of-covariance

(regression) estimation model to examine the effect of Lone Star AP CSP on each of these survey-based outcomes.

Slightly different sets of Lone Star AP CSP teachers were included in each year's analysis based on which teachers completed surveys and taught AP CSP each program year as well as in the pre-program baseline year. Separately each year, the eligible Lone Star AP CSP teachers were matched with similar non-program AP CSP teachers who also completed surveys. The numbers of teachers included in each year's analysis are as follows:

- Year 1: 9 Lone Star teachers and 54 matched comparison teachers
- Year 2: 7 Lone Star teachers and 49 matched comparison teachers
- Year 3: 7 Lone Star teachers and 35 matched comparison teachers

The Lone Star teachers were matched with comparison teachers on the baseline measures of the outcomes. On each outcome in each year, the baseline mean difference between the Lone Star teachers and the selected comparison teachers was less than 0.25 standard deviations. Analytic models controlled for the baseline measure of the outcome and for selected teacher characteristics, such as the number of years the teacher had held a full-time job in computer science, that were found to be predictive of the baseline measure among the teachers in the sample.

The key findings from the analysis of impact on instructional practice are presented in Exhibit ES.4 and can be summarized as follows:

- After one year of implementation, the Lone Star AP CSP program had a significant positive effect on the extent to which teachers encouraged students to use cognitive problem-solving strategies.
- After two years of implementation, the program had a significant and large positive effect on each of the three measures of instructional practice we examined.
- After three years of implementation, the program had no effect on any of the three measures of instructional practice.

Exhibit ES.4. Lone Star AP CSP Impact on Instructional Practices: Hedges' *g* Effect Sizes

Outcome Measure	Year 1 (2020–21)	Year 2 (2021–22)	Year 3 (2022–23)
Engaging Students in Practices of Computer Science	0.53	1.36***	0.07
Encouragement of Cognitive Problem-Solving Strategies	0.76*	1.00**	0.02
Use of Pedagogically Based Learning Strategies	0.38	1.31***	0.23

***$p < .001$, **$p < .01$, *$p < .05$, †$p < .10$.
Note. Positive effect sizes indicate effects in the desired direction according to the theory of action.

In summary, we found that the Lone Star AP CSP program had positive effects on participating teachers' instructional practices after the first and second years of implementation but not after the third. A number of caveats should be kept in mind in interpretation of the results. Perhaps most importantly, the measures of teachers' instructional practice are based on the teachers' own self-report on retrospective year-end surveys, and it is possible that teachers may have misreported or misrepresented their instructional practices (whether intentionally or not). Other limitations include very small samples (which changed from year to year) and a lack of generalizability.

Impact on Student AP CSP Performance and Participation (RQs 4–6)

As detailed in the final column of the program logic model, the ultimate aims of the Lone Star AP CSP program were to increase the likelihood that (a) AP CSP course takers succeed in the course, and (b) high school students take the AP CSP course and exam. This section examines whether the program achieved these aims during its first, second, and third years of operation.

The outcome measures for the student impact study fall into two categories: performance and participation. Performance pertains to how well students did in AP CSP, which was theorized to be affected by Lone Star AP CSP mainly through the supports for teachers and the supports for students, and participation pertains to how many students took the AP CSP course and exam, which was theorized to be affected mainly by the supports for schools.

We examined four school-level outcomes in the analysis of the impact of the Lone Star AP CSP program on student AP CSP performance and participation:

- the percentage of AP CSP exam takers earning a 3, 4, or 5 on the exam *(key measure of student performance)*;
- the percentage of students in grades 9–12 earning a 3, 4, or 5 on the AP CSP exam *(supplementary measure #1 of student performance)*;

- the percentage of students in grades 9–12 earning a 1 or 2 on the AP CSP exam *(supplementary measure #2 of student performance)*; and
- the percentage of students in grades 9–12 taking the AP CSP exam *(measure of student participation)*.

Separately in each of the three outcome years, the student impact study used a quasi-experimental matched comparison group design and an analysis-of-covariance (regression) estimation model to examine the effect of Lone Star AP CSP on the performance and participation outcomes, followed by meta-analysis to combine the estimates from two samples of schools (those that had AP CSP exam takers in the baseline year and those that did not).

Slightly different sets of Lone Star AP CSP schools were included in each year's analysis based on data availability, and they were matched with similar comparison schools separately each year. The numbers of schools included in each year's analysis are as follows:

- Year 1: 28 Lone Star schools and 187 matched comparison schools
- Year 2: 34 Lone Star schools and 185 matched comparison schools
- Year 3: 32 Lone Star schools and 216 matched comparison schools

Lone Star AP CSP schools were matched with comparison schools on the baseline measures of the outcomes (or, for schools that did not have AP CSP exam takers in the baseline year, proxy measures involving all AP subjects), with the same set of comparison schools across all four outcomes. For all samples and outcomes, the baseline mean differences between the Lone Star schools and the selected comparison schools were less than 0.25 standard deviations. Analytic models controlled for the baseline measure of the outcome; total school enrollment in the outcome year; percentages of students that were black, Hispanic, female, and economically disadvantaged; charter school status; and locale type (city, suburb, or town or rural).

The key findings from the analysis of impact on AP CSP exam performance and participation are presented in Exhibit ES.5 and can be summarized as follows:

- After one year of implementation:
  - The Lone Star AP CSP program did not have a significant effect on schools' percentages of AP CSP exam takers that received a qualifying score on the exam.
  - The program also did not have a significant effect on schools' percentages of students in grades 9–12 that received a qualifying score—or a nonqualifying score—on the AP CSP exam.
  - There was also no significant effect on schools' percentages of students in grades 9–12 that took the AP CSP exam.

- After two years of implementation:
  - The Lone Star AP CSP program had a borderline-significant negative effect on schools' percentages of AP CSP exam takers that received a qualifying score on the exam.
  - The program had a significant negative effect on schools' percentages of students in grades 9–12 that took the AP CSP exam.
  - Likely because of the negative effect on the percentages of students taking the exam, the program also had a significant negative effect on schools' percentages of students that received a qualifying score on the AP CSP exam, as well as a borderline-significant negative effect on schools' percentages of students that received a *non*qualifying score.
- After three years of implementation:
  - The Lone Star AP CSP program did not have a significant effect on schools' percentage of AP CSP exam takers that received a qualifying score on the exam.
  - The program also did not have a significant effect on schools' percentages of students in grades 9–12 that received a qualifying score on the AP CSP exam, but it did have a significant effect on schools' percentages of students in grades 9–12 that received a *non*qualifying score. This effect was negative, which is in the desired direction.
  - The program also had a significant negative effect on schools' percentages of students in grades 9–12 that took the AP CSP exam.

Exhibit ES.5. Lone Star AP CSP Impact on AP CSP Exam Performance and Participation: Hedges' *g* Effect Sizes

Outcome Measure	Year 1 (2020–21)	Year 2 (2021–22)	Year 3 (2022–23)
Percentage of AP CSP exam takers earning a 3, 4, or 5 on the exam	-0.11	-0.27†	-0.07
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the AP CSP exam	0.13	-0.49**	-0.13[a]
Percentage of students in grades 9–12 earning a 1 or 2 on the AP CSP exam	0.14	-0.23†	-0.50**
Percentage of students in grades 9–12 taking the AP CSP exam	0.10	-0.40**	-0.42*

***$p < .001$, **$p < .01$, *$p < .05$, †$p < .10$.

[a] indicates significant ($p < .05$) heterogeneity between the two subsamples.

Note. Effect sizes are from meta-analyses combining two subsamples of schools. Positive effect sizes indicate effects in the desired direction according to the theory of action, except for the third row outcome (percentage of students in grades 9–12 earning a 1 or 2 on the AP CSP exam), where negative effect sizes are in the desired direction.

In summary:

- In none of the three years did the program have a significant positive impact on the percentage of AP CSP exam takers that received a qualifying score on the exam.
- In none of the three years did the program have a significant positive impact on schools' percentages of students in grades 9–12 that took the AP CSP exam. To the contrary, in Years 2 and 3 the program's effect on student participation in the AP CSP exam was significantly negative.

Conclusion

In comparison to matched comparison schools and controlling for baseline measures and school demographics, schools in the Lone Star AP CSP did not display higher rates of success on or participation in the AP CSP exam in any of the three program years. Therefore, the program did not have its theorized positive effects either on AP CSP exam performance or participation. Based on survey self-report from a small number of Lone Star AP CSP teachers and matched comparison teachers, the program appears to have had a positive impact on AP CSP instructional practices in Year 2 (and to a lesser extent in Year 1), but the self-reported basis for these measures and the very small sample size, in combination with the null findings of program impact on students' AP CSP exam performance, call this instructional practice finding into question. Relatively low levels of program implementation and uptake probably help explain why the program did not achieve its hypothesized positive impacts on students. Lingering effects of the COVID-19 pandemic may have hampered implementation efforts.

It is also worth bearing in mind for the AP performance and participation findings that the comparison schools included in those analyses could have been implementing other interventions pertaining to AP CSP. There have been numerous such interventions devised over the past several years. Against such a backdrop, significant program effects become all the more difficult to detect.

Chapter 1. Introduction

This report describes the independent evaluation of the Lone Star Advanced Placement (AP) Computer Science Principles (CSP) program and presents the evaluation's findings.

The Lone Star AP CSP program, funded by an Education and Innovation Research (EIR) early-phase grant from the U.S. Department of Education, was developed by the National Math + Science Initiative (NMSI) to increase student access to and success in the AP CSP course in Texas high schools. The American Institutes for Research® (AIR®) conducted the evaluation of the program, including studies of implementation, impact on teachers' instruction, and impact on students' participation in and performance on the AP CSP exam.

Overview of the Lone Star AP CSP Program

The Lone Star AP CSP program began in summer 2020. It was designed as a three-year program whose key components were teacher professional development and supports for students and schools. Chapter 2 describes these components in detail and includes a logic model (Exhibit 2.1) depicting how they were theorized to bring about the desired outcomes. As shown in the logic model, the teacher professional development component was theorized to improve teachers' content knowledge and instructional practice related to AP CSP, which in turn was theorized to improve students' success in the AP CSP course; the supports for students were theorized to directly improve students' success in the AP CSP course as well as increase the likelihood that they would take the course and the AP exam; finally, the supports for schools were designed to encourage schools to increase their AP CSP offerings and eliminate barriers to taking AP CSP, which in turn were theorized to improve students' likelihood of taking the course.

The three-year duration was an important feature of the proposed program, particularly in terms of the supports for teachers. The initial EIR proposal (NMSI, 2019) makes several mentions of a three-year training arc, sequence, or cycle for teachers, proposing the use of "a sustained teacher professional development [PD] cycle which responds to teachers' learning progressions over time" (p. 11). In fact, teacher professional development sequenced over three years was what distinguished the Lone Star program from NMSI's regular and well-established College Readiness Program (CRP) and therefore made it innovative for EIR purposes. However, given teacher turnover and other challenges resulting from the COVID-19 pandemic, NMSI was not able to ensure that the initially recruited teachers would participate for the full three years. Looking to provide support for as many teachers as possible, NMSI permitted additional teachers at the participating schools to join the program in its second and third years.

Thirty-seven schools, all in Texas (by design), were recruited by NMSI in spring 2020 to participate in the program. Over the three years of the program, a total of 61 AP CSP teachers participated, 27 of whom participated for all three years.

ABOUT AP CSP

Advanced Placement® Computer Science Principles was one of the College Board's 38 AP courses offered during the study period. Launched in the 2016–17 school year, AP CSP is a relatively new AP course; it was designed to attract students traditionally underrepresented in computer science—particularly Black, Hispanic, and female students (College Board, n.d.-a). According to the fall 2020 *Course and Exam Description* (which was in effect throughout the program period, although it has since been updated):

> *AP Computer Science Principles introduces students to the breadth of the field of computer science. In this course, students will learn to design and evaluate solutions and to apply computer science to solve problems through the development of algorithms and programs. They will incorporate abstraction into programs and use data to discover new knowledge. Students will also explain how computing innovations and computing systems, including the Internet, work, explore their potential impacts, and contribute to a computing culture that is collaborative and ethical. (College Board, 2020, p. 7)*

The document goes on to note that the course does not have a designated programming language and is "equivalent to a first-semester, college-level breadth course in computer science." Moreover, although no prior experience with computer science is required, completion of a first-year high school algebra course is recommended as a prerequisite. The course is built around five "big ideas": creative development, data, algorithms and programming, computing systems and networks, and impact of computing. There are also six cross-cutting computational thinking practices and skills: computational solution design, algorithms and program development, abstraction in program development, code analysis, computing innovations, and responsible computing.

Overview of the Evaluation

AIR conducted the independent evaluation of the Lone Star AP CSP program. The evaluation examined the implementation of the program, including the extent to which it was implemented with fidelity to its intended design, teachers' and students' levels of participation in the program, and participating teachers' perceptions of the program. In addition, the evaluation examined the impact of the program on teachers' instructional practice and on

student participation in and performance on the AP CSP exam. Both of these impact studies employed a quasi-experimental design (QED) using matched comparison groups, with matched teachers from nonparticipating schools for the teacher impact study and matched schools for the student impact study (i.e., a cluster QED). Both employed an analysis-of-covariance (regression) estimation strategy to estimate program effects. Separate analyses were conducted for program impacts after one year, two years, and three years of program implementation.

Research Questions

Using administrative and survey data, the evaluation addressed the following research questions:

1. Was the Lone Star AP CSP program implemented with fidelity in
    a. its first year (the 2020–21 school year)?
    b. its second year (the 2021–22 school year)?
    c. its third year (the 2023–23 school year)?
2. To what extent did Lone Star AP CSP program participants make use of the program and find it useful?
3. Did participation in the Lone Star AP CSP program improve AP CSP teachers' instructional practices in the AP CSP course, compared to AP CSP teachers in the business-as-usual condition, after one year, after two years, and after three years?
4. What was the impact of one year of participation in the Lone Star AP CSP program, compared to business-as-usual AP CSP participation, on high schools' rates of (a) AP CSP exam takers receiving a qualifying score on the AP CSP exam, and (b) students taking the AP CSP exam?
5. What was the impact of *two* years of participation in the Lone Star AP CSP program, compared to business-as-usual AP CSP participation, on the same outcomes?
6. What was the impact of *three* years of participation in the Lone Star AP CSP program, compared to business-as-usual AP CSP participation, on the same outcomes?

Interim Reports

Prior to this final report, the evaluation produced the following interim reports:

> Holtzman, D. J. (2021). *Independent evaluation of Lone Star AP CSP: Report on teacher and student use of the program in year 1.* American Institutes for Research.

Holtzman, D. J. (2023, February). *Independent evaluation of Lone Star AP CSP: Impact on instructional practices after the first and second years of program implementation.* American Institutes for Research.

Jones, K. T., & Holtzman, D. J. (2023, April*). Independent evaluation of Lone Star AP CSP: Implementation and uptake in the first and second years of the program.* American Institutes for Research.

These reports are available upon request to the study authors.

Independence of the Evaluation

The evaluation of the program was completed *independently* by AIR. NMSI provided the administrative data AIR used to assess fidelity of implementation and also played a small role in procuring some of the data used in the teacher impact study, as shall be detailed in the relevant chapters of this report. However, AIR alone processed and analyzed all data and independently conducted all other associated evaluation activities (e.g., obtaining the data for the student impact study and selecting comparison teachers and schools). NMSI was invited to provide feedback on a draft of this report, but AIR retained sole editorial control.

Preregistration of the Study Design

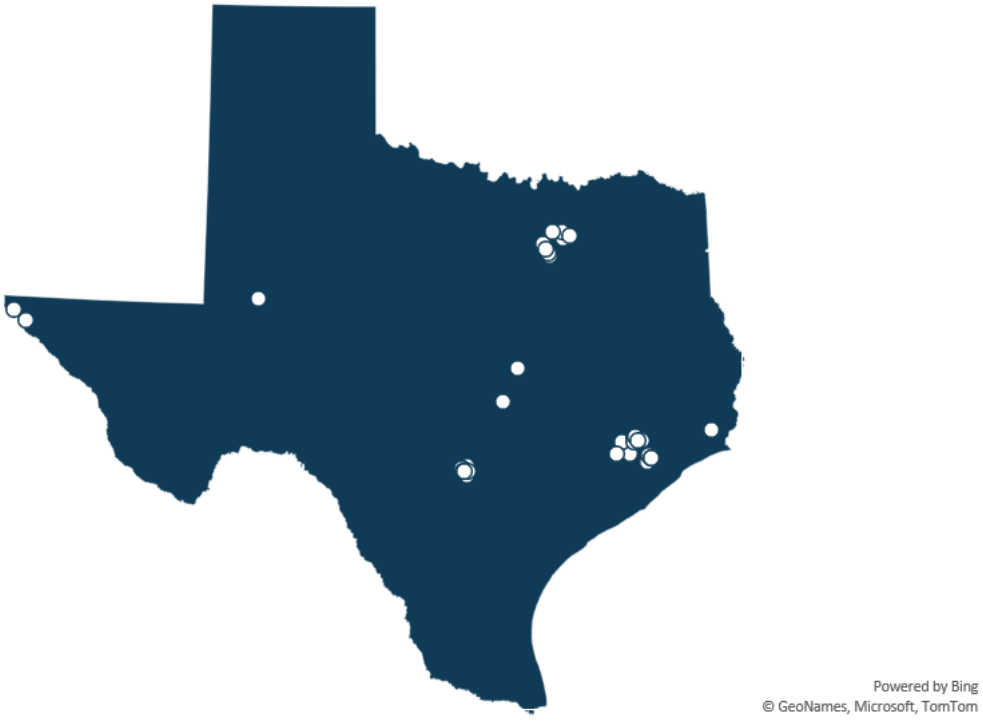
The impact study design was preregistered in the Registry of Efficacy and Effectiveness Studies (REES) under registry ID 9040, with separate subregistries for the teacher and student impact studies, in December 2021. No major changes to the study design were made following the preregistration, although there were some minor changes, particularly pertaining to the method used to select comparison schools in the student impact study and the numbers of teachers and schools included in both impact studies. In addition, for the student impact study, we decided to conduct separate analyses for different matching blocks (based on whether schools had AP CSP exam takers prior to the implementation period), and we slightly revised the list of covariates included in the regression models.

Overview of the Intervention Group Sample (Schools and Teachers)

School Sample

As already noted, 37 Texas schools were recruited by NMSI to participate in the Lone Star AP CSP program; these schools constitute the school-level intervention-group sample for the program. According to data from the National Center for Education Statistics (NCES), the 37 schools belonged to 15 local education agencies (LEAs), with the number of schools per LEA ranging from one to seven, and averaging 2.5. Exhibit 1.1 shows the approximate locations of the schools.

Exhibit 1.1. Locations of the Lone Star AP CSP Schools



Over half of the schools (23) were classified by NCES as being in cities; 12 are suburban and two are rural. Sixteen (43%) are charter schools.

Across the 37 schools, the average enrollment in the 2019–20 (baseline) school year was 1,489 students. On average, 46% of students were female, and 68% qualified for free or reduced-price lunch. About two thirds (66%) were Hispanic, 16% were Black, 10% were White, and 7% were Asian. Appendix Table A-1 provides further detail about these statistics.

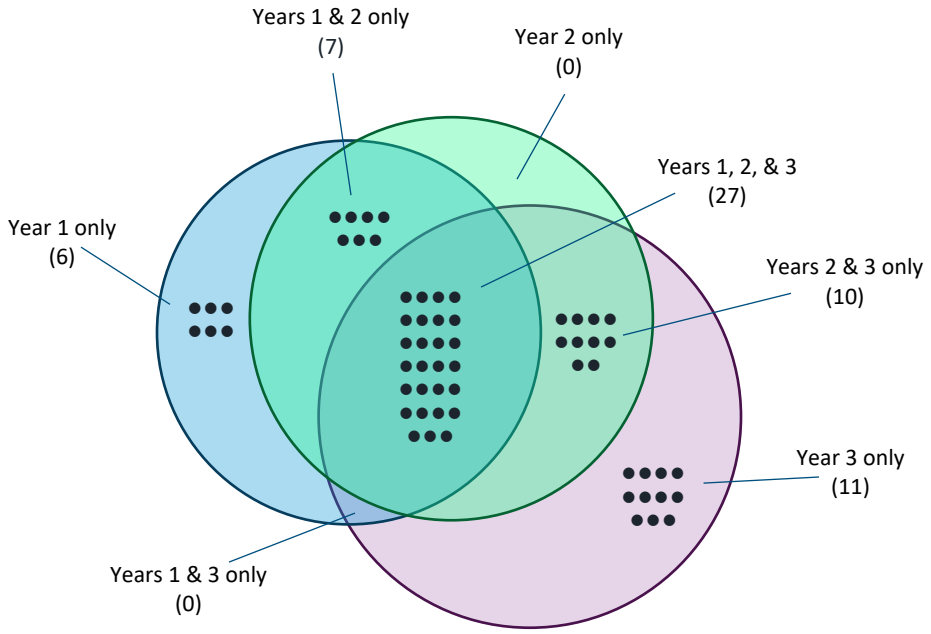
Characteristics of the subset of Lone Star schools included in the student impact study for each program year, as well as of their matched comparison schools, will be described in Chapters 5, 6, and 7 and their associated appendices.

Teacher Sample

In the program's first year (the 2020–21 school year), there were 40 AP CSP teachers in the 37 participating schools; these 40 teachers constitute the Year 1 program sample for teacher-focused analyses. Subsequently, the participating schools experienced some turnover in the teachers of the AP CSP course, due at least in part to lingering effects of the COVID-19 pandemic, and more turnover than the program had initially expected given its intended focus on teachers' learning progressions. Specifically, by the end of the second year (the 2021–22

school year), six of the initial 40 teachers had left the program (by virtue of leaving their school[1]) and 10 new teachers had joined, resulting in a sample of 44 teachers for Year 2. Then, between Years 2 and 3, seven more teachers left and 11 new teachers joined, resulting in a sample of 48 teachers for Year 3. Exhibit 1.2 shows the overlap among the three years' teacher samples.

Exhibit 1.2. Overlap Among the Three Years' Teacher Samples



Only a subset of the teachers could be included in the teacher impact study; this is discussed in Chapter 4, which also provides background information about the Lone Star teachers and comparison teachers constituting the teacher impact study sample.

Organization of This Report

The following six chapters of this report address the six research questions, in order with one question per chapter. A final conclusion chapter summarizes the main findings for each research question and then briefly synthesizes them.

[1] NMSI did not track teachers' departures from schools, so we used survey invitation email bounces as a proxy (see Chapter 3 for more information about the surveys). We assumed any teacher whose survey invitation (sent in May of each school year) did *not* bounce had been at the participating Lone Star AP CSP school that year and was therefore still part of the program sample for that school year, regardless of whether they had actively participated in any of the program's offerings.

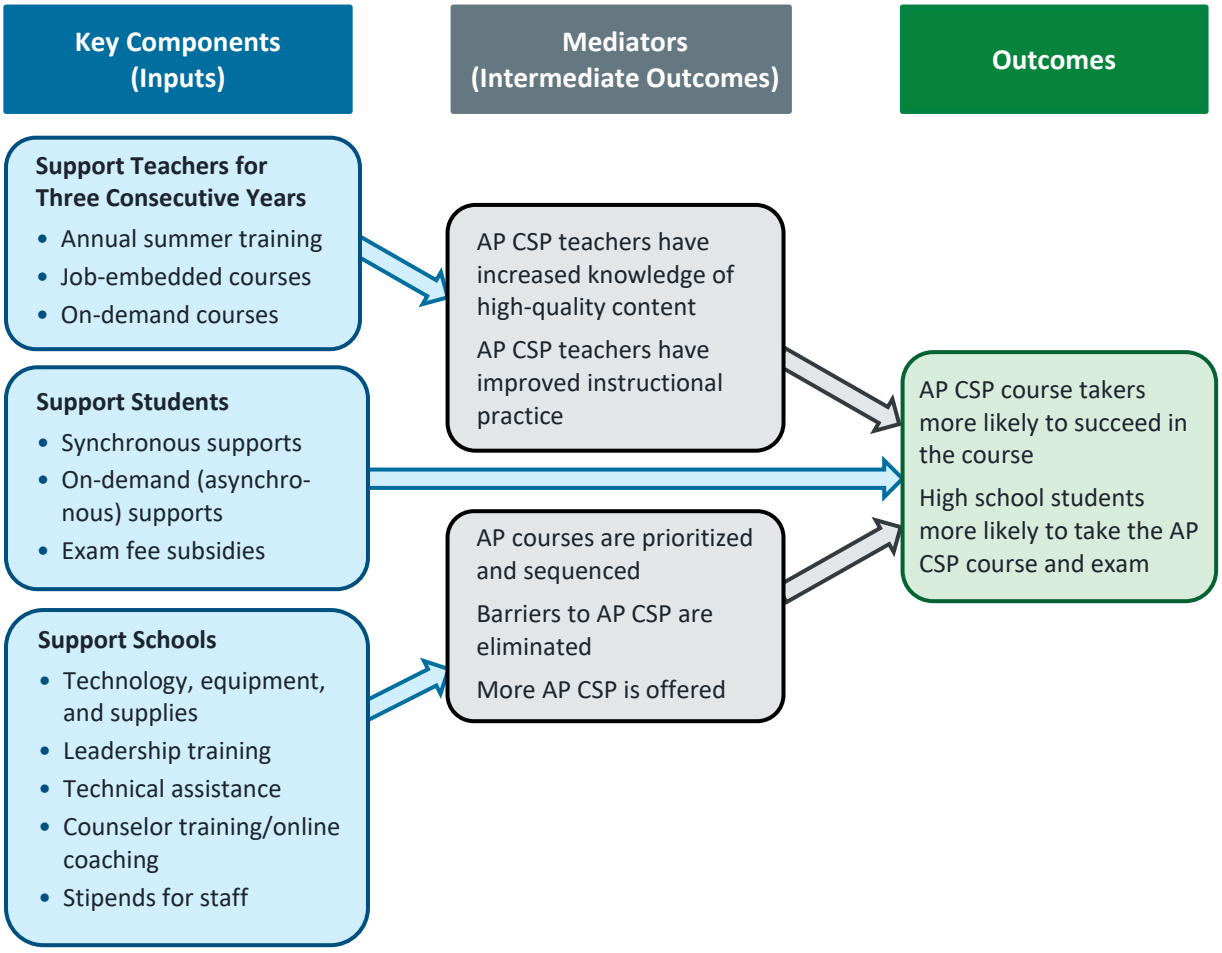
Chapter 2. Fidelity of Implementation

This chapter examines the extent to which the Lone Star AP CSP program was implemented with fidelity in each of its three years. The research question addressed is as follows:

1. Was the Lone Star AP CSP program implemented with fidelity in
    a. its first year (the 2020–21 school year)?
    b. its second year (the 2021–22 school year)?
    c. its third year (the 2023–23 school year)?

As noted in Chapter 1, the key components of the Lone Star AP CSP program were teacher professional development and supports for students and schools. Exhibit 2.1 provides further detail about these components and illustrates how they were theorized to bring about both intermediate and longer-term outcomes. Additional detail is provided throughout this chapter.

Exhibit 2.1. Lone Star AP CSP Logic Model



Our approach to measuring fidelity, which involved close collaboration with NMSI, had the following steps:

1. Identify and describe the main elements constituting each of the three key components in each of the program's three years.
2. Identify a data source by which implementation of each element could be gauged and develop a plan for the collection of the needed data.
3. Determine how to score implementation of each element and set a threshold for "adequate" implementation of that element.
4. Set a threshold for adequate implementation of each of the key components based on the individual element scores.
5. Collect the data and score each element and component.

KEY FINDINGS

- The teacher supports component was not implemented with fidelity in any of the three years.
- The student supports component was implemented with fidelity in all three years.
- The school supports component was implemented with fidelity in the first year but not in the second or third years.

Steps 1 through 4, which occurred relatively early on in the evaluation of Lone Star AP CSP, are summarized in implementation matrices provided in the evaluation's design plan, submitted in August 2021 and approved by the Abt Associates EIR Evaluation Team the following month. These matrices, with only minor revisions following their initial development and approval, are presented (along with the overall fidelity determinations for each component in each year) in Appendix Exhibit B-1.

Because essentially all of the program elements were delivered directly from NMSI to the participants, it made sense to measure fidelity of implementation at the program level. Accordingly, the fidelity measures cover the entire sample of schools and teachers participating in Lone Star AP CSP, and most of the data used in our fidelity calculations were administrative data provided by NMSI. We measured fidelity separately for each of the three years of the program.

We now discuss each of the three program components in turn, providing further information about what the component consisted of, how we gauged fidelity of implementation (FoI), and what we found.

Teacher Supports

Lone Star AP CSP's primary teacher supports were three sets of professional development offerings:

- Annual summer training: a multi-day set of workshops and courses offered in July preceding each school year called the NMSI Summer Series and, starting in 2021, a two-day Content Workshop held in June (like the Summer Series, in preparation for the following school year). All three years' Summer Series were held virtually;[1] the Content Workshop that constituted the first activity of Year 2 was also held virtually, but the Content Workshop constituting the first activity of Year 3 was held in person.[2] Through direct trainings, practicums, expert speaker presentations, and panels, the summer offerings focused on topics such as computer science content, computer science pedagogical content knowledge, the AP CSP exam, STEM instructional practices, general teaching pedagogy, lesson planning and delivery, instructional decision making, and equitable STEM pathways.
- Job-embedded courses: courses offered virtually (but synchronously) during the school year. Several of these courses were meant to respond to teachers' immediate needs based on where they were in their teaching of the AP CSP course and, as such, were called "just in time" courses (i.e., offered at an opportune time to be maximally useful in the teaching of the course). There were also courses called "funds of knowledge" and "mock exam." Starting in Year 2, the job-embedded courses were concentrated into fall and spring sessions.
- On-demand courses: a variety of online courses, offered asynchronously for teachers to take at their convenience and based on their own needs. Individual courses focused on specific computer science topics (e.g., algorithms and developing programs, data mining, impact of computing on society) as well as topics related to teaching more generally

ADDITIONAL TEACHER SUPPORTS

NMSI also offered the Lone Star AP CSP teachers opportunities to participate in two online learning communities: one called the Teacher Villages and the other a Facebook group. We considered including provision of these learning communities as an element in the FoI matrix, but NMSI, after internal deliberation, indicated to the evaluation team that they should be excluded because they were not a key element of the teacher supports component. Chapter 3 presents exploratory findings on teachers' participation in these communities.

[1] The Summer Series was originally designed as an in-person event, but the COVID-19 pandemic necessitated that it be moved online. In 2020 and 2021 (Years 1 and 2), the Summer Series lasted two weeks; in 2022 (Year 3), it was one week.
[2] The Year 3 (2022) Content Workshop was held in person twice, once in June and once in July. Teachers could choose to attend one or the other. The Content Workshop sessions were also recorded and made available virtually (asynchronously), along with assignments, for teachers who did not attend in person.

(e.g., "being a culturally responsive educator," "family and community engagement") and the AP exams.

Measurement of Fidelity

The measurement of FoI for the summer and job-embedded courses differed somewhat across the three years. In Years 1 and 2, NMSI provided AIR with data files showing the number of summer and job-embedded credits earned by each Lone Star teacher. By Year 3, however, NMSI had moved away from a credit accounting metric, and simply tracked attendance. NMSI and AIR set thresholds for the summer training element to be considered implemented with fidelity, as follows:

- Year 1: 80% of teachers needed to have earned at least 16 credits
- Year 2: 70% of teachers needed to have earned at least 16 credits
- Year 3: 65% of teachers needed to have attended the summer training

For the job-embedded courses, meanwhile, we set the following thresholds:

- Year 1: 60% of teachers needed to have earned at least 10 credits
- Year 2: 60% of teachers needed to have earned at least 10 credits
- Year 3: 60% of teachers needed to have attended both the fall and spring trainings

In Years 2 and 3, we employed a weighting scheme in the calculation of the percentages of teachers; see the box titled "Weighting for Elements 1 and 2 of the Teacher Supports Component" for the details.

WEIGHTING FOR ELEMENTS 1 AND 2 OF THE TEACHER SUPPORTS COMPONENT

As noted in Chapter 1, the Lone Star AP CSP program was intended as a three-year program for a single cohort of teachers to continuously improve their teaching practice over the whole course of the program. Given that continuity of participation over time was what made the program distinctive and innovative (as compared to NMSI's well-established College Readiness Program), we thought it was important to consider the continuity in the fidelity determinations. However, likely due to the COVID-19 pandemic and its associated challenges, the participating schools experienced more churn than NMSI had expected in the teachers of the AP CSP course. Exhibit 1.2 in Chapter 1 illustrates the amount of turnover from year to year.

In an effort to remain true to the original "three-year training arc for a single cohort of teachers" intent of the program without omitting the newly joining teachers entirely, we decided to include newly joining teachers in the Year 2 and Year 3 fidelity determinations but to give more weight to

continuing teachers, especially those continuing from Year 1 (i.e., the teachers, who, through a sequenced course of three-year learning and professional growth, were theorized to have the greatest impact on students by the final year of the program). This weighting applies to the two program elements calculated based on individual teacher data: the summer training and the job-embedded courses.

For the Year 2 fidelity determination, we assigned a weight of two to each of the 34 teachers continuing from Year 1 and a weight of one to each of the 10 teachers new to the program in Year 2—i.e., one unit of weight for each year of participation in the program. This meant that the denominator for the Year 2 percentage calculations was 78: (34 continuing teachers × 2 years) + (10 new teachers × 1 year).

For the Year 3 fidelity determination, we again assigned one unit of weight for each year of participation in the program, as follows:

- a weight of three to each of the 27 teachers continuing from Year 1
- a weight of two to each of the 10 teachers who started in Year 2 and continued into Year 3
- a weight of one to each of the 11 teachers who started in Year 3

Accordingly, the denominator for the Year 3 percentage calculations was 112: (27 teachers from Year 1 × 3 years) + (10 teachers from Year 2 × 2 years) + (11 new teachers × 1 year).

The on-demand courses—the third element of the teacher support component of the program—were optional for participating teachers. Accordingly, AIR and NMSI agreed that this program element could be considered implemented with fidelity as long as at least five such courses were *available* to teachers each year, regardless of how many teachers took them or how many of the courses any given teacher took. NMSI created an account for AIR to log in to the Blackboard platform through which the courses were offered and confirm that at least five courses were available.

To consider the overall teacher supports component to have been implemented with fidelity in each year, we set as the criterion that at least two of the three elements (summer training, job-embedded courses, on-demand courses) had to have been implemented with fidelity, based on the criteria described in the preceding paragraphs.

Results

The teacher supports component was not implemented with fidelity in any of the three years. In all three years, the on-demand supports element met the threshold for fidelity, but neither of the other two components—summer training and job-embedded courses—met their respective thresholds. Thus, in each year the teacher supports component received an overall score of only 1 (out of 3), instead of the 2 required for fidelity (Exhibit 2.2).

For the Year 1 (2020) summer training, 34 of the 40 teachers (85%) registered, 32 (80%) attended, and 24 (60%) earned at least one course credit, but only 13 (32.5%) completed at least 16 credits, falling considerably short of the pre-established 80% threshold. For the job-embedded courses, only 14 teachers (35%) registered for at least one course, and only seven (17.5%) received any credits. Although all seven of these teachers received at least 10 credits (the qualifying number of credits), the fidelity criteria specified that at least 60% of teachers would need this many credits.

For the Year 2 (2021) summer training, only 40% of teachers attended the summer training, and even fewer—just 28%—earned the qualifying 16 credits, again falling far short of the 70% threshold.[3] Similarly, for the job-embedded courses, fewer than half (47%) of the teachers registered for at least one course, and only 17% received 10 or more credits. This, too, falls considerably short of the established 60% threshold. Exhibit B-2 in the appendix provides both the unweighted (raw) and weighted number of teachers upon which these percentage calculations are based.[4]

In Year 3 (2022), 46% of teachers attended the summer training, again well below the threshold (65%).[5] Similarly, for the job-embedded courses, only 43% of the teachers participated in both the fall and spring sessions, while the criteria called for 60% doing so. Even if the criteria had called for 60% of teachers attending *either* the fall or spring sessions (rather than both), it still would not have been met, as only 49% did so. Exhibit B-3 in the appendix provides both the unweighted (raw) and weighted number of teachers upon which these percentage calculations are based.[6]

[3] All percentages in this paragraph are based on weighted counts, as described in the box earlier.
[4] Unweighted, 25% of teachers (11 of 44) earned 16 or more credits in the summer training, and 16% of teachers (7 of 44) earned 10 or more credits in the job-embedded courses. Therefore, the fidelity criteria for these two elements still would not be met without the weighting. In fact, the weighted percentages are higher than the unweighted percentages.
[5] All percentages in this paragraph are based on weighted counts, as described in the box earlier.
[6] Unweighted, 48% of teachers (23 of 48) attended the summer training, and 48% of teachers (23 of 48) attended both spring and fall sessions of the job-embedded courses. Therefore, as in Year 2, the fidelity criteria for these two elements still would not be met without the weighting.

Exhibit 2.2. Fidelity of Implementation for Teacher Supports Component, by Element

Teacher supports			Year 1 results (2020–21 school year)			Year 2 results (2021–22 school year)			Year 3 results (2022–23 school year)		
Element	Description	Threshold for fidelity of implementation	Sample for fidelity measure	Finding	Fidelity met?	Sample for fidelity measure	Finding	Fidelity met?	Sample for fidelity measure	Finding	Fidelity met?
Summer training	Professional development offered during summer (Summer Series preceding all three years; Content Workshop preceding Years 2 and 3)	Y1: At least 80% of teachers received 16+ credits of summer training Y2: At least 70% of teachers received 16+ credits of summer training Y3: At least 65% of teachers attended summer training	40 teachers (full program sample)	33% (13 of 40)	No	44 teachers (78 weighted) (full program sample)	28% (22 of 78)[a]	No	48 teachers (112 weighted) (full program sample)	46% (51 of 112)[c]	No
Job-embedded courses	Synchronous virtual courses offered during the school year	Y1 & Y2: At least 60% of teachers received 10 credits Y3: At least 60% of teachers attended fall and spring sessions	40 teachers (full program sample)	18% (7 of 40)	No	44 teachers (78 weighted) (full program sample)	17% (13 of 78)[b]	No	48 teachers (112 weighted) (full program sample)	43% (48 of 112)[d]	No
On-demand courses	Asynchronous virtual courses	At least 5 asynchro-nous virtual courses were available	Whole Lone Star AP CSP program	20 courses	Yes	Whole Lone Star AP CSP program	20 courses	Yes	Whole Lone Star AP CSP program	12 courses	Yes
Number of elements meeting FoI criteria					**1**			**1**			**1**

[a] 11 of 34 teachers continuing from Year 1, and 0 of 10 teachers new in Year 2

[b] 6 of 34 teachers continuing from Year 1, and 1 of 10 teachers new in Year 2

[c] 11 of 27 teachers continuing from Year 1, 6 of 10 teachers continuing from Year 2, and 6 of 11 teachers new in Year 3

[d] 10 of 27 teachers continuing from Year 1, 5 of 10 teachers continuing from Year 2, and 8 of 11 teachers new in Year 3

Student Supports

The student supports component of the Lone Star AP CSP program had three main elements:

- Synchronous supports: NMSI held live virtual study sessions, collectively covering 12 topics, aimed at AP CSP students.
- On-demand supports: Students had access to an online library of modules and video recordings relevant to the AP CSP course. In addition, NMSI coaches periodically held office hours for students to drop in on for additional support.
- AP exam cost subsidies: NMSI offered to cover at least some of the cost of student registration fees for the AP CSP exam. The exact value was dependent on students' free and reduced-price lunch status.

Measurement of Fidelity

As with the on-demand courses for teachers, participation in the synchronous study sessions and the on-demand supports was optional for students, so NMSI needed only to document that these supports were available to be considered to have implemented them with fidelity. (See Chapter 3 for details about uptake of the student resources.) To be considered to have implemented the synchronous sessions element with fidelity, NMSI needed to provide at least one study session per topic; for the on-demand supports element, NMSI needed to have offered at least one online module and one office hour session per topic. As documentation, NMSI provided AIR with a schedule of the AP CSP synchronous sessions and office hours, as well as a list of the other on-demand supports that were available in each of the three years. In Years 1 and 2, NMSI also provided a login for AIR to confirm that these resources were available.

The AP exam subsidies were also in some sense optional. Each year, NMSI notified the participating schools that the subsidies were available, but NMSI only provided the subsidies to the schools that explicitly requested them following the notification. To document that the subsidies were offered, NMSI provided AIR with the notifications they sent to the participating schools. NMSI also shared documentation that each school that requested the subsidy received the full amount. To be considered to have implemented this element with fidelity, every school that requested subsidies had to have received them.

To consider the overall student supports component to have been implemented with fidelity in each year, we set as the criterion that all three elements (synchronous supports, on-demand supports, and AP exam subsidies) had to have been implemented with fidelity, per the criteria described in the preceding paragraphs.

Results

In Year 1, all three of the student supports elements received scores of 100% and therefore easily met their thresholds for fidelity, as did the overall component. Twelve study sessions and 12 online modules were offered to students, as were office hours, thereby satisfying the fidelity criteria for the first two elements. For the third element, only one school requested AP exam fee subsidies, and it received the funds, so the criterion for this element was also met.[7]

The data provided by NMSI indicated that the student supports were also implemented with fidelity in Years 2 and 3. As in Year 1, all three elements received scores of 100%, the only difference being that two schools requested and received AP exam fee subsidies.

See Exhibit 2.3 for a detailed breakdown of these results.

[7] We do not know for sure why so few schools requested the AP exam subsidies, but it may be because both the College Board and the state of Texas also help cover the cost of AP exams for eligible students: see https://tea.texas.gov/about-tea/news-and-multimedia/correspondence/taa-letters/texas-advanced-placement-ap/international-baccalaureate-ib-incentive-program-state-exam-fee-subsidy-for-the-2022-2023-school-year.

Exhibit 2.3. Fidelity of Implementation for Student Supports Component, by Element

Student supports			Year 1 results (2020–21 school year)			Year 2 results (2021–22 school year)			Year 3 results (2022–23 school year)		
Element	Description	Threshold for fidelity of implementation	Sample for fidelity measure	Finding	Fidelity met?	Sample for fidelity measure	Finding	Fidelity met?	Sample for fidelity measure	Finding	Fidelity met?
Synchronous supports	Live virtual study sessions split between 12 topics; optional for students	At least one study session per topic will be available to all students	Whole Lone Star AP CSP program	12 sessions	Yes	Whole Lone Star AP CSP program	12 sessions	Yes	Whole Lone Star AP CSP program	12 sessions	Yes
On-demand (asynchronous) supports	Online library of modules; video recordings from NMSI coaches with office hours available; optional for students	At least one online module and one office hour session per topic will be available to students	Whole Lone Star AP CSP program	12 modules, each with office hours	Yes	Whole Lone Star AP CSP program	12 modules, each with office hours	Yes	Whole Lone Star AP CSP program	12 modules, each with office hours	Yes
Exam fee subsidies	Students are given a subsidy for the AP exam fee (value dependent on the student's free and reduced-price lunch status)	All schools that request subsidies receive their requested amount	Whole Lone Star AP CSP program	1 school requested subsidies; it received them	Yes	Whole Lone Star AP CSP program	2 schools requested subsidies; both received them	Yes	Whole Lone Star AP CSP program	2 schools requested subsidies; both received them	Yes
Number of elements meeting FoI criteria					**3**			**3**			**3**

School Supports

NMSI offered a number of supports to the 37 schools participating in the Lone Star AP CSP program. Some of these supports were offered every year, while others were only offered in specific years. These supports were as follows:

- Technology, equipment, and supplies: NMSI offered the schools laptops and/or tablets for students to use in AP CSP classrooms. This support was optional and only offered at the outset of the program (i.e., Year 1), with the assumption that any equipment provided would last all three years.
- Leadership training: School leaders were invited to attend the same Summer Series as the teachers (see the section above for details about this offering). Leaders could choose to participate in any of the three years but were required to attend only in one of them.
- Counselor training: Starting in Year 2, NMSI staff offered coaching sessions, conducted virtually but synchronously, to the counselors at the Lone Star schools. In Year 2, plans called for seven 1-hour coaching sessions. For Year 3, NMSI decreased the frequency of sessions to only three but increased the duration of each session to 2 to 3.5 hours.
- Technical assistance: NMSI was to hold regular meetings with the program director at each school, the site coordinator, and school counselors.
- Stipends for school staff: The Lone Star program included stipends, up to $2,000 per year, for staff members at each participating school as follows:
  - Director: $500 per year for overseeing program implementation, providing data to NMSI, and following up on requests
  - Site coordinator: $500 per year for ensuring that teachers and students had access to supports
  - Counselor: $1,000 per year for engaging with and coaching students and for implementing practices supporting the recruitment of students, especially underrepresented students, into AP CSP

Measurement of Fidelity

As with the AP exam subsidies for students, NMSI notified participating schools that laptops and tablets were available. Schools then had to request the equipment, and if a school did not request equipment, none was provided. To be considered to have implemented this element with fidelity, NMSI needed to provide equipment to all of the schools that requested it. Because equipment was only provided in Year 1, fidelity of implementation for this component is applicable only for that year. To document that the equipment was offered, NMSI provided AIR with the notifications they sent to all participating schools. NMSI also shared shipping records to document that the schools that requested equipment received it.

For the leadership training, NMSI provided AIR with data on the number of leaders from each school who attended the training in each of the three summers (2020, 2021, and 2022). Because school leaders were required to attend the training in only one of the three years, this element was not included in the fidelity determination until Year 3. The criterion for implementation with fidelity was that at least 75% of the schools needed to have a school leader attend the training in at least one of the three years.

The training and coaching for school counselors was not offered in Year 1 and was therefore excluded from that year's fidelity analysis. For Years 2 and 3, NMSI provided attendance records for the sessions held. For determining fidelity of implementation, we set as the criterion that at least 80% of the schools had to have at least one counselor attend at least 5 hours of the training, defined as at least five of the 1-hour coaching calls in Year 2 and at least two of three (longer) sessions in Year 3.[8]

To measure implementation of the technical assistance element, AIR requested documentation of NMSI's contacts with schools—specifically, a list of the dates of meetings with each school. To be considered to have implemented this element with fidelity, NMSI needed to show that at least 80% of schools met regularly with NMSI program managers over the course of each year. For the purposes of this measure, we defined "regularly" as at least five meetings during the school year, with at least one meeting in the fall (August through November), at least one in winter (December through February) and at least one in spring (March through May).

For the measurement of FoI for the school stipends, NMSI agreed to provide a list showing the stipend dollar amount provided to each school. For this element to have been implemented with fidelity, at least 80% of schools needed to receive the full $2,000 in stipends.

Because of the variation across the three years in which school support elements were offered, the determination of FoI for the overall school supports component differed somewhat each year. In Year 1, when there were three elements (equipment, technical assistance, and school staff stipends), we set as the criterion that at least two of the three elements had to have been implemented with fidelity, per the criteria described in the preceding paragraphs. Year 2 also had three elements (counselor training, technical assistance, and stipends), so again the criterion was that two of the three elements had to be offered and documented. Year 3, however, had four elements (leadership training, counselor training, technical assistance, and stipends); accordingly, we set the criterion that three of them had to be offered and documented for the component to be implemented with fidelity.

[8] In Year 3, one counseling session was 3.5 hours, another was 3 hours, and the third was 2 hours. Therefore, attending any two of the three sessions would constitute at least 5 hours.

Results

In Year 1, two of the three school support elements (equipment and staff stipends) met the threshold for fidelity, while one (technical assistance) did not. For equipment, 23 schools requested equipment, and all 23 schools received it (100%). For school stipends, 33 schools (89%) received at least $2,000 in stipends, surpassing the 80% threshold.

NMSI did not collect detailed information on the provision of technical assistance during the 2020–21 school year and therefore was not able to provide AIR with any records of their meetings with schools. Because we could not determine how many schools received the technical assistance, we were compelled to give the Lone Star program a score of 0 on that indicator in Year 1. Even so, given the requirement that only two of the three school supports elements had to be implemented with fidelity, NMSI met the fidelity requirement for the overall school supports program component in Year 1.

In Year 2, on the other hand, the school supports component was *not* implemented with fidelity, with only one of the three elements—school staff stipends—meeting the fidelity criteria. For technical assistance, all 37 schools met with NMSI at least five times, and all had meetings in the winter and spring quarters, but only 76% (28 of 37) met with NMSI in the fall quarter. Therefore, NMSI just missed the 80% threshold for implementation with fidelity—had one more school met with NMSI in the fall, the threshold would have been reached. Counselor training and coaching, meanwhile, fell much further short of the mark, with only 14% (five of 37) attending at least five sessions, as compared to a threshold of 80%. The one element that did meet the 80% threshold in Year 2 was school staff stipends, with 81% (30 of 37) schools receiving at least $2,000 in stipends.

The school supports component was again *not* implemented with fidelity in Year 3. Of the four component elements applicable in Year 3, only one—the training for school leaders—met the fidelity criteria. For this element, 89% of the schools (33 of 37) sent a leader to the training in at least one program year, easily surpassing the 75% threshold. (In fact, 25 schools, or 68%, sent a leader to multiple years' trainings.) But the other three elements all fell far short. For technical assistance, only three schools (8%) met the criteria for meeting with NMSI, whereas the threshold was 80% of schools.[9] For counselor training, which also had an 80% threshold, only 32% of the schools (12 of 37) met the criteria.[10] Finally, only 22 schools (59%) received the program's stipends for school staff, once again compared to a threshold of 80% of schools.

Exhibit 2.4 presents additional detail on these results.

[9] One additional school met with NMSI five times over the course of the 2022–23 school year, but none of the meetings were in the spring period. Across all 37 schools, the average number of technical assistance meetings was two. Nine schools (24%) had no meetings at all, so even if the criterion had been a single meeting during the year, it still would not have been met.
[10] Fourteen of the schools (39%) had no counselor training during Year 3, according to NMSI's records.

Exhibit 2.4. Fidelity of Implementation for School Supports Component, by Element

School supports			Year 1 results (2020–21 school year)			Year 2 results (2021–22 school year)			Year 3 results (2022–23 school year)		
Element	Description	Threshold for fidelity of implementation	Sample for fidelity measure	Finding	Fidelity met?	Sample for fidelity measure	Finding	Fidelity met?	Sample for fidelity measure	Finding	Fidelity met?
Technology, equipment, and supplies	Schools receive laptops and/or tablets	Y1: 100% of schools that requested equipment received it from NMSI Y2 & Y3: Not applicable	23 schools (the rest did not request equip-ment)	100% (23 of 23)	Yes	Not applicable in Year 2			Not applicable in Year 3		
Leadership training	Ongoing training for school leadership	Year 3 (cumulative for all 3 years): At least 75% of schools have a leader who participates in at least one summer	Not applicable in Year 1			Not applicable in Year 2			37 schools (full program sample)	89% (33 of 37)	Yes
Technical assistance	All programmatic support from NMSI, including regular meetings with the school director, site coordinator, and a counselor	At least 80% of schools meet with NMSI at least five times, including once in fall (Aug–Nov), once in winter (Dec–Feb), and once in spring (Mar–May)	37 schools (full program sample)	0%	No	37 schools (full program sample)	76% (28 of 37)	No	37 schools (full program sample)	8% (3 of 37)	No

(table continues on next page)

Exhibit 2.4. Fidelity of Implementation for School Supports Component, by Element, continued

School supports			Year 1 results (2020–21 school year)			Year 2 results (2021–22 school year)			Year 3 results (2022–23 school year)		
Element	Description	Threshold for fidelity of implementation	Sample for fidelity measure	Finding	Fidelity met?	Sample for fidelity measure	Finding	Fidelity met?	Sample for fidelity measure	Finding	Fidelity met?
Counselor training/ online coaching	Seven 1-hour coaching calls with Google Classroom support	Y1: Not applicable Y2: 80% of schools have at least one counselor attend 5 or more coaching calls Y3: 80% of schools have at least one counselor attend 2 or more coaching calls (5 or more hours)	Not applicable in Year 1			37 schools (full program sample)	14% (5 of 37)	No	37 schools (full program sample)	32% (12 of 37)	No
Stipends for school staff	Monetary incentive for school director ($500), site coordinator ($500), and counselor ($1,000)	At least 80% of schools receive stipends totaling $2,000	37 schools (full program sample)	89% (33 of 37)	Yes	37 schools (full program sample)	81% (30 of 37)	Yes	37 schools (full program sample)	59% (22 of 37)	No
Number of elements meeting FoI criteria					**2**			**1**			**1**

Summary and Discussion

In summary, we found that in Year 1 (the 2020–21 school year) the student and school supports were implemented with fidelity, but the teacher supports were not. In each of Years 2 and 3, only the student supports component was implemented with fidelity; the other two components—teacher supports and school supports—were not. (See Exhibit 2.5.)

Exhibit 2.5. Summary of FoI, by Component

Component	Year 1 (2020–21)	Year 2 (2021–22)	Year 3 (2022–23)
Teacher supports	Not met	Not met	Not met
Student supports	Met	Met	Met
School supports	Met	Not met	Not met
Total number of components met	**2 of 3**	**1 of 3**	**1 of 3**

In none of the three years, then, were all components implemented with fidelity, and in each of Years 2 and 3, only one of the three components was implemented with fidelity. Moreover, in none of the years was the teacher supports component implemented with fidelity. Given the emphasis on this component in the program's original EIR proposal (as the program element that, as a three-year arc, made the Lone Star AP CSP program innovative and different from NMSI's regular College Readiness Program) and the prominent role it plays in the program's theory of action for improving student performance on the AP CSP exam, this finding is of serious concern.

The COVID-19 pandemic may understandably have interfered with program implementation. Although one might think that the pandemic's interference would mainly have been limited to Year 1 (i.e., the first full year after the onset of the pandemic), the educational landscape continued to shift in the ensuing years as a result of pandemic-caused upheaval, thereby creating continued challenges for program implementation.

The overall low levels of implementation fidelity, particularly in Years 2 and 3, must be kept in mind in interpreting the impact results that will be discussed in Chapters 4 through 7. Given the implementation findings, it would not be surprising to find minimal or no impact. That said, it is also worth considering what participating teachers themselves reported to be their program uptake levels and how these teachers viewed the program and its elements. This is the subject of the next chapter.

Chapter 3. Participant Engagement and Perceptions

In May of each of the three years of implementation of the Lone Star AP CSP program, NMSI invited the Lone Star AP CSP teachers to complete a survey developed and administered by AIR. Each year's survey, conducted online, asked the teachers about their participation in the program activities over the course of the preceding school year, as well as their perceptions of usefulness of each activity they participated in and barriers to participation.[1] The surveys also asked about the participation of the teachers' AP CSP students, given that the program includes supports and resources for students along with professional development and supports for teachers. In addition, the final year's survey included a set of three-year reflection questions for the subset of teachers who had participated in the program in all three years. This chapter summarizes the teachers' responses to these questions, with the overarching research question being as follows:

2. To what extent did Lone Star AP CSP program participants make use of the program and find it useful?

KEY FINDINGS

- Based on teachers' self-report on surveys, teacher participation in Lone Star AP CSP activities was relatively low in Year 1 and somewhat higher in Year 2. There was little change from Year 2 to Year 3 in participation levels, but perceptions of usefulness of the teacher activities increased across the three years.
- Use of the student resources, as reported by teachers, increased from year to year, as did teacher perceptions of the usefulness of these resources.
- Lack of awareness about the program's supports and resources was a commonly cited reason for nonuse of particular activities in all three years, but more so in Year 1 than Year 2, and more so in Year 2 than Year 3.
- At the close of the program in Year 3, most teachers (60%) reported that overall, they had found the program to be at least moderately useful.
- The subset of teachers who participated in all three years of the program had mixed opinions about the cohesiveness of the program across the three years, such as how well it implemented the envisioned three-year arc of professional development.

Whereas Chapter 2 presented findings related to the "official" determination of fidelity of implementation of Lone Star AP CSP, this chapter presents descriptive findings about the extent to which teachers used various program elements and found them useful; there were no pre-specified thresholds or criteria against which to judge

[1] The survey also included questions about teachers' instructional practice that are being used in the evaluation's impact study (see the following chapter). Comparison teachers recruited and selected by AIR also completed the instructional practices section of the survey but were routed out of the section on use of the Lone Star program.

implementation. The findings in this chapter are also exploratory, in that they are based on teachers' self-report to surveys that, for better or worse, were optional; thus, as shall be noted in various places, we cannot be sure that the Lone Star AP CSP teachers who completed the survey for a given year represented the full sample of teachers constituting that year's program sample.[2] Conceivably, respondents may have been more engaged with the program than nonrespondents, in which case reported participation levels, and perhaps satisfaction levels as well, may be higher than actual levels. Nevertheless, findings presented in this chapter can, along with the findings presented in the previous chapter, help provide context for the impact findings reported in the subsequent chapters.

Survey Administration and Response

As noted in Chapter 1, 40 teachers participated in Lone Star AP CSP in Year 1 (2020–21), 44 participated in Year 2 (2021–22), and 48 participated in Year 3 (2022–23). A total of 27 teachers participated in all three years.

For the 2021 survey (corresponding to the end of the first implementation year), we received responses from 31 of the 40 teachers in the Year 1 Lone Star teacher sample (78% response rate). One of the respondents, however, reported at the beginning of the survey that they had not taught AP CSP in the 2020–21 school year, making them ineligible to complete the rest of the survey. Of the remaining 30 teachers, 28 finished the survey and answered at least some of the questions in the section about participation in the Lone Star AP CSP program.

For the 2022 survey, we received responses from 32 of the 44 teachers constituting the Year 2 Lone Star sample (73% response rate). Twenty-five of the 32 reported teaching AP CSP in the 2021–22 school year; the other seven said they had not.[3] Of the 25 teachers who said they had, all 25 answered at least some of the questions about their Lone Star AP CSP participation in 2021–22.

For the 2023 survey, we received responses from 36 of the 48 teachers constituting the Year 3 Lone Star sample (75% response rate). Thirty-three of the 36 reported teaching AP CSP in the 2022–23 school year; the other three said they had not.[4] Of the 33 teachers who said they had, 30 answered at least some of the questions about their program participation in 2022–23.

[2] The voluntary nature of the surveys, the lack of 100% response, and the resulting uncertainty about the representativeness of respondents are the reasons why the survey data were not considered as a data source for our fidelity of implementation determinations.

[3] These seven teachers, all of whom had been in the Year 1 sample, were from seven different schools. Four of the seven teachers indicated that their school had other AP CSP teachers in the 2021–22 school year, but NMSI's records showed another Lone Star teacher at only one of these four schools.

[4] These three teachers, all of whom had been in the Year 1 sample, were from three different schools. Two of the three schools had other AP CSP teachers in 2022–23, based both on NMSI's records and other survey responses.

The remainder of this chapter summarizes the responses of the 28 Year 1 teachers (i.e., 70% of the Year 1 sample), the 25 Year 2 teachers (57% of the Year 2 sample), and the 30 Year 3 teachers (63% of the Year 3 sample) who reported on their program participation. A total of 12 teachers answered the questions in all three years' surveys (i.e., 44% of the 27 teachers who were in all three years' samples); on occasion, we will report results for this smaller group to provide an alternate look at change over the three years. It should be noted that the Year 2 and Year 3 teacher weighting described in Chapter 2 is *not* applied in this chapter, given this chapter's more exploratory intent.

Finally, 18 teachers who were in all three years' program samples (though not necessarily all three years' *survey* samples) answered the 2023 survey's set of questions about their perceptions of the whole three-year arc of the Lone Star AP CSP program. Their responses to these questions will be summarized toward the end of the chapter.

Teacher Participation in the Main Components of Lone Star AP CSP

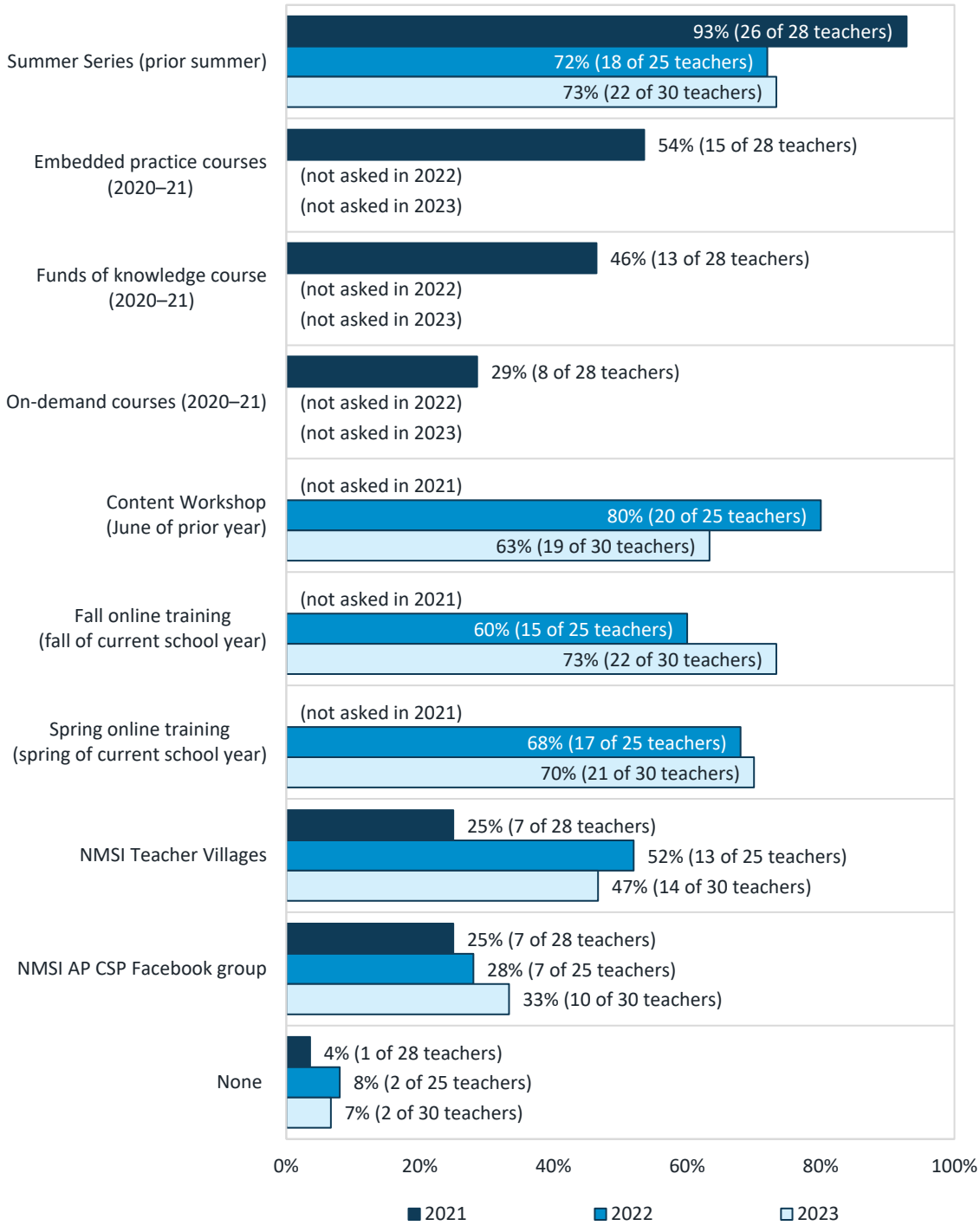
All three years' surveys asked Lone Star teachers to indicate which of several program activities they had participated in during the preceding school year, including the previous summer. However, as a result of changes to the program's activities and emphases, the list of activities changed somewhat after the first (2021) survey. Exhibit 3.1 specifies which activities were asked about on each survey.

Exhibit 3.1. Lone Star AP CSP Activities Listed on the Three Years' Teacher Surveys

Lone Star activities listed in all three surveys	Lone Star activities listed in 2021 survey only	Lone Star activities listed in 2022 and 2023 surveys only
• NMSI Summer Series in the previous summer • NMSI Teacher Villages (online learning communities) • NMSI AP CSP Facebook group	• Embedded practice ("just in time") courses • Funds of knowledge course • On-demand courses	• Content Workshop for AP CSP (online in June of the previous calendar year) • Fall online training (just in time course or funds of knowledge course) in fall of the current school year • Spring online training (just in time course or mock exam course) in spring of the current school year

In Year 1, all but one of the 28 teachers reported having participated in at least one of the six activities, with the average being 2.7 of the six activities. As shown by the top, dark-colored bars in Exhibit 3.2, far and away the most commonly reported activity was the Summer Series, which 93% of the teachers reported participating in. About half of the teachers reported participating in embedded practice courses (54%) and the funds of knowledge course (46%). About one in four teachers reported participating in each of the three other activities listed in the 2021 survey: on-demand courses, NMSI Teacher Villages, and the NMSI AP CSP Facebook group.

Exhibit 3.2. Teacher Participation in Lone Star AP CSP Activities, by Year



In Year 2 (middle bars in Exhibit 3.2), the percentage of teachers who reported participating in the Summer Series dropped somewhat to 72%, and there were two teachers who reported not having participated in any activities, as opposed to only one in Year 1. However, participation in

most of the other activities appears to have increased. In particular, the percentage of teachers reporting participation in the Teacher Villages doubled, from 25% to 52%. In addition, although the courses and trainings were asked about differently on the two years' surveys, a majority of 2022 survey respondents indicated participating in each of the three trainings (60% to 80%), compared to about half who reported participating in the roughly equivalent options on the 2021 survey (embedded practice and funds of knowledge courses). Of the six activities, the average number that teachers participated in was 3.6.

Year 3 fairly closely resembled Year 2. The average number of activities was again 3.6, and again, two teachers reported not having participated in any activities. For particular activities (lower, light-colored bars in Exhibit 3.2), the Year 3 percentage of teachers participating was higher than the Year 2 percentage for some (especially the fall online training) and lower for others (especially the Content Workshop), but all in all, the Year 3 percentages are not very different from those in Year 2.

Among the 12 teachers for whom we have data for all three years, the average number of activities (of six total) they participated in increased from 3.5 in Year 1 to 4.9 in Year 2, then dropped back down slightly to 4.6 in Year 3. As suggested by these averages, nearly all of the 12 teachers participated in more activities in Year 2 than in Year 1, and then either had the same number of activities or one fewer in Year 3.

Therefore, in multiple respects, teacher participation in Lone Star AP CSP activities appears to have increased from Year 1 to Year 2, and then remained relatively constant in Year 3. However, as noted in the introduction to this chapter, we cannot rule out some degree of bias in findings on participation levels, based on which teachers completed the teacher survey.

Perceptions of Usefulness of the Main Activities

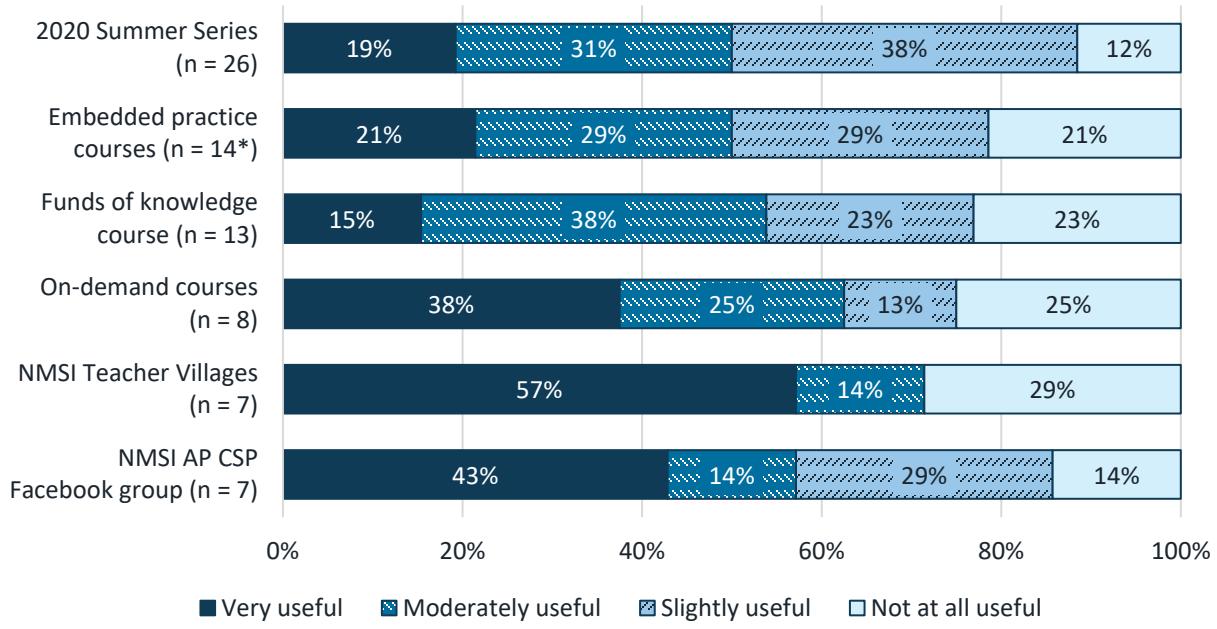
In all three years, teachers who reported participating in any given Lone Star AP CSP activity were asked how useful they had found that activity, on a 4-point scale from "not at all useful" (coded as 1) to "very useful" (coded as 4). As the first graph in Exhibit 3.3 shows, the 2020 Summer Series, the embedded practice courses, and the funds of knowledge course were rated as at least moderately useful by about half of the teachers who had reported participating in each activity. Higher percentages of participants rated the other three activities as at least moderately useful—up to 71% for the NMSI Teacher Villages. However, as discussed earlier (and also as shown by the *n* sizes in parentheses on the bar labels in Exhibit 3.3), fewer teachers participated in these three activities. In addition, although the Teacher Villages received the highest percentage of teachers rating the activity as at least moderately useful (71%)—and as very useful (57%)—it also had the highest percentage of teachers rating it as not at all useful

(29%). Across activities, the average usefulness rating among the Year 1 teachers (*n* = 27 who rated the usefulness of at least one activity) was 2.7 (below *moderately useful*).

In Year 2, the four training activities (Summer Series, Content Workshop, and fall and spring online trainings) were each rated as at least moderately useful by 60% to 70% of participants (second graph in Exhibit 3.3), suggesting an increase in perceived usefulness of these activity types. On the other hand, percentages rating the Teacher Villages and Facebook group as at least moderately useful dropped somewhat from Year 1 to Year 2, as did the percentages rating these activities as very useful. For the Teacher Villages, the drop in usefulness ratings is despite—or possibly related to—the larger number of teachers participating in this activity. Among the 23 Year 2 teachers who rated the usefulness of at least one activity, the average usefulness rating across activities was 2.9 (very close to *moderately useful*).

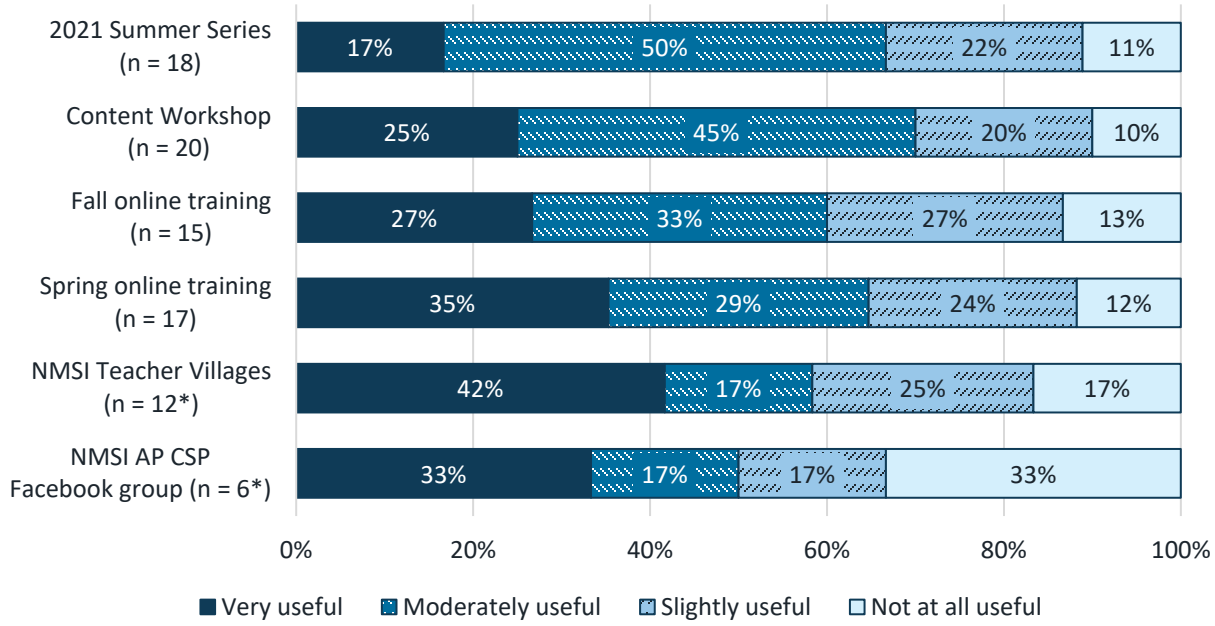
Exhibit 3.3. Teacher Ratings of Usefulness of Lone Star AP CSP Teacher Activities

a. Year 1 (2021 survey, about 2020–21 program activities)



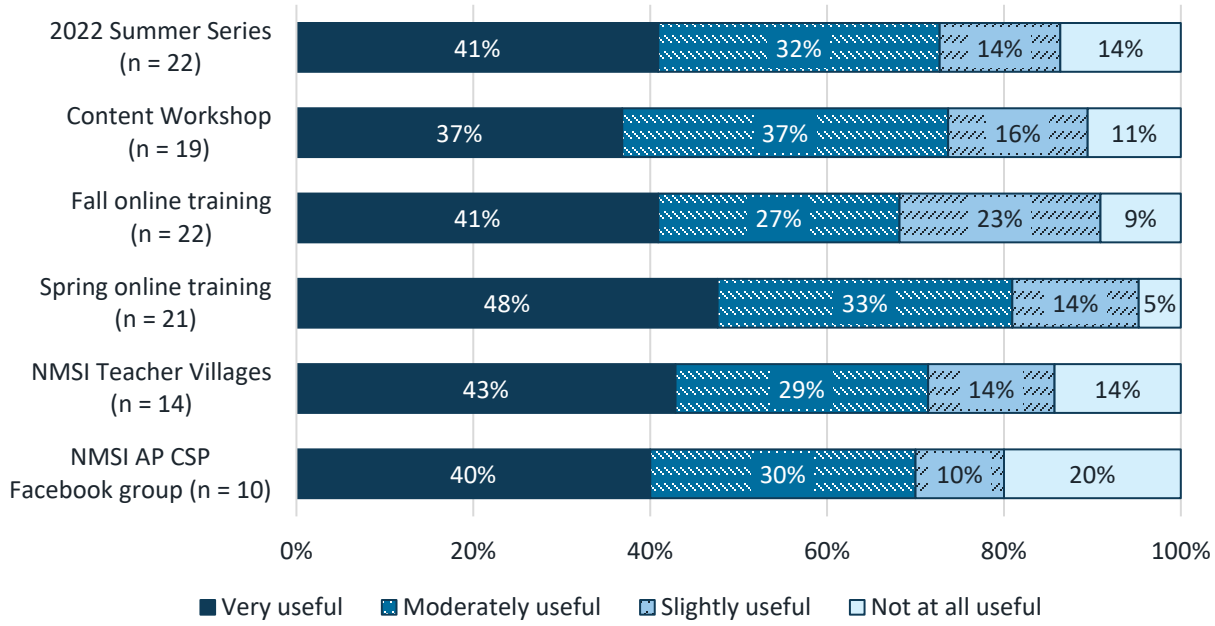
*One teacher who reported participating in this activity did not rate its usefulness.

b. Year 2 (2022 survey, about 2021–22 program activities)



*One teacher who reported participating in this activity did not rate its usefulness.

c. Year 3 (2023 survey, about 2022–23 program activities)



Compared to Year 2 respondents, Year 3 respondents were substantially more likely to rate activities as very useful (dark blue portions of bars in the third graph in Exhibit 3.3), and somewhat more likely to rate activities as at least moderately useful (first two bar segments).

The increases from Year 2 to Year 3 in the percentage of respondents rating at least moderately useful are particularly notable for the spring online training, the Teacher Villages, and the Facebook group. Even so, the cross-activity average usefulness rating among Year 3 teachers (*n* = 28 who rated the usefulness of at least one activity) was 2.9, the same as the average rating among the Year 2 teachers.[5]

Reasons for Nonparticipation

For each activity that teachers did *not* report participating in, they were asked why not. The survey provided a list of six to eight options (with the exact number varying depending on the activity), as well as an "other" option with space to write in a reason. Some of the listed options were the same across all of the activities, whereas others were customized to the particular activity.

In Year 1, the most common response teachers selected for each activity was that *they did not know about it*. This response was checked by 80% of the nonparticipants (12 of 15) for the funds of knowledge course, and by about half of the nonparticipants for each of the other five activities, ranging from 43% (nine of 21) for the Facebook group up to 57% (12 of 21) for the NMSI Teacher Villages. The next most common reason respondents gave for nonparticipation in each activity (with the exception of the 2020 Summer Series[6]) was that they did not have time. This was particularly the case for the Teacher Villages and the Facebook group, for which 48% and 38% of respondents, respectively, indicated a lack of time as a reason for their nonparticipation.

In Year 2, not knowing about an activity was still a fairly common reason given for nonparticipation, especially for the spring online training, which five of the eight nonparticipants (63%) reported not knowing about. Not knowing about the activity was also the most common reason given for nonparticipation in the Teacher Villages (seven of 12 nonparticipants, i.e., 58%), the Facebook group (nine of 18 nonparticipants, i.e., 50%), and the fall online training (four of 10 nonparticipants, i.e., 40%). For the June 2021 Content Workshop and the 2021 Summer Series, the most common reason teachers gave for nonparticipation was that they had not yet been hired to their school; this reason was given by three of five

[5] Comparing Exhibit 3.3b to 3.3c, it may seem surprising that the average usefulness rating was 2.9 in both years. This is because the usefulness rating was calculated at the teacher level. That is, we calculated an average rating for each teacher, across the activities that the teacher participated in and rated; the overall average is the average (across teachers) of the individual teacher average ratings. If instead we calculate an average rating at the *activity* level—that is, across teachers, the average usefulness rating of each activity—and then average *those* ratings across the six activities, the Year 2 average is 2.8 and the Year 3 average is 3.0. (In more technical terms, the issue is that of row means versus column means: the mean of row means and the mean of column means will be equal only if all cells in the table are nonmissing. That is not the case here because not every teacher rated the usefulness of all six activities.)

[6] Of the two teachers who did not report participating in the 2020 Summer Series, one said that they did not know about it, and the other said that they had not yet joined the program.

nonparticipants (60%) for the Content Workshop and four of seven nonparticipants (57%) for the Summer Series.[7]

In Year 3, not having time for the activity and not knowing about it tended to be the top-reported reasons for nonparticipation in most of the activities. Not having time was the most commonly selected reason for nonparticipation in the 2022 Summer Series (marked by three of eight nonparticipants, tied with the number who marked not having yet joined the Lone Star AP CSP program), the fall online training (marked by four of eight nonparticipants, tied with the number who marked not knowing about it), the spring online training (marked by five of nine), and the Teacher Villages (12 of 16). Not knowing about the activity, meanwhile, was the most commonly selected response for nonparticipation in the June 2022 Content Workshop (marked by three of 11 nonparticipants), the fall online training (four of eight, tied with not having time), and the Facebook group (eight of 20). Even among the 18 survey respondents who had been in the Lone Star AP CSP program since Year 1, five of them indicated not knowing about at least one activity offered in Year 3.

Lone Star AP CSP Resources for Students

As discussed in Chapter 2, Lone Star AP CSP also includes resources directed toward students: live online study sessions, on-demand student supports, and online office hours with NMSI coaches. We did not solicit feedback directly from students on these resources (which, as noted in Chapter 2, were optional), but we did include questions about the student resources on the survey for teachers.

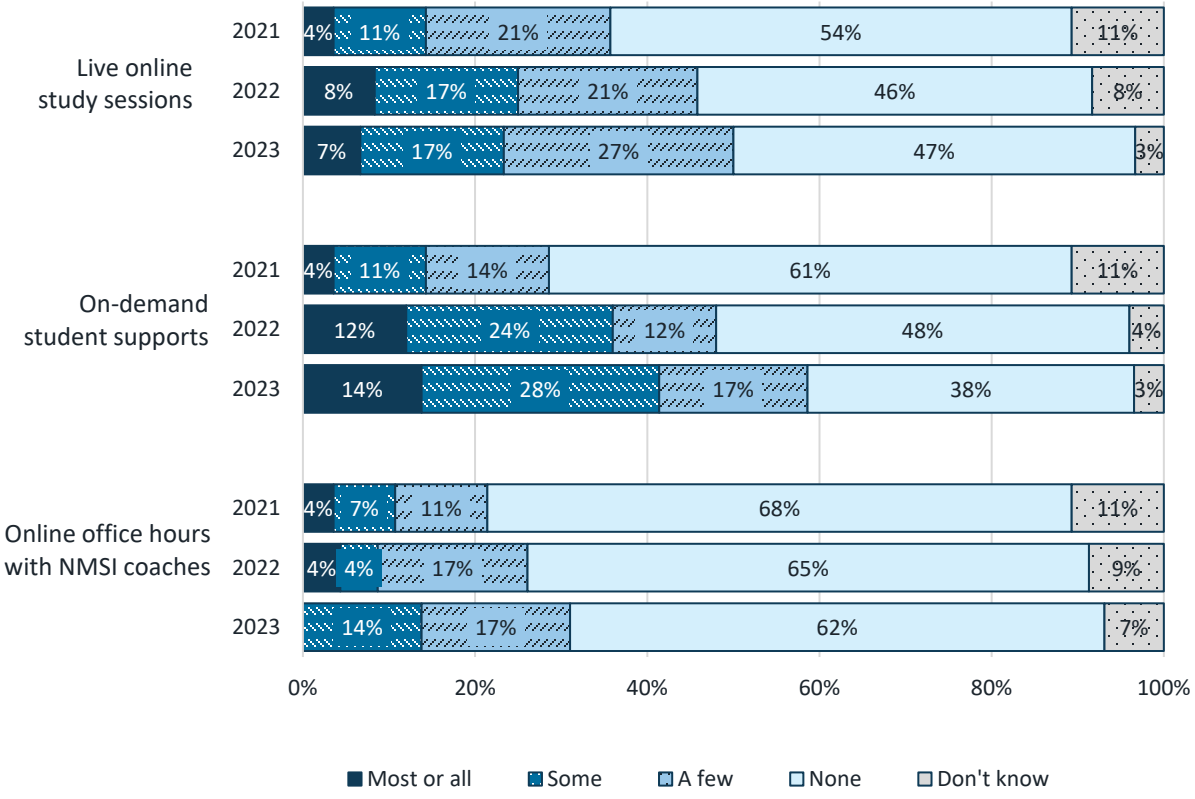
Level of Use of the Resources for Students

Use of the student resources, as reported by teachers, increased from year to year (Exhibit 3.4). In Year 1, only 10 of the 28 teachers (36%) reported any student use of any of the three student resources. For each of the three student resources individually in Year 1 (top bar of each group in Exhibit 3.4), more than half the teachers indicated that *none* of their students had used the resource. The most widely used student resource, according to teachers, was the live online study sessions, which 36% of teachers reported that at least a few students had used. The least-used resource was the online office hours, which only 21% of teachers reported that at least a few students had used. For none of the three resources did more than a single teacher (4%) report that *most or all* students used the resource in Year 1—and it was the same teacher who

[7] One of the four teachers who gave this reason for not participating in the 2021 Summer Series *did* indicate participating in the June 2021 Content Workshop, which preceded the 2021 Summer Series. Moreover, this teacher was *not* new to the program in Year 2. Therefore, this teacher either misunderstood the 2021 Summer Series nonparticipation question or mistakenly marked that they had not yet been hired to their school. This teacher also indicated two other reasons for not participating in the 2021 Summer Series.

gave that response for all three resources. There were also three teachers (11%) who said they did not know how many students used any of the resources.

Exhibit 3.4. Teacher Report on How Many Students Used the Student Resources, by Year



Note. In Year 1 (2021), *n* = 28 for all three bars. In Year 2 (2022), *n*s for the three bars are 24, 25, and 23 respectively (from top to bottom). In Year 3 (2023), *n* = 30 for the bar for live online study sessions and *n* = 29 for the bars for the other two resources.

In Year 2, 14 teachers (56%) reported any student use of the three resources, up from 10 teachers (36%) in Year 1. Almost half of the teachers (46%) reported that at least a few students had used the live online study sessions in Year 2 (middle bar of top group in Exhibit 3.4), compared with only 36% in Year 1. Similarly, for the on-demand student supports, 48% of teachers reported use by at least a few students in Year 2, compared with only 29% in Year 1. Moreover, for both of these resources, the largest increases were for "most or all students" and "some students" as opposed to only "a few students." Most notably, for the on-demand supports, 36% of the Year 2 teachers reported that more than a few students had accessed the resource; this exceeds the 29% who reported use by *any* students in Year 1. And three teachers, up from just one in Year 1, reported that most or all students had used one or more of the three resources in Year 2.

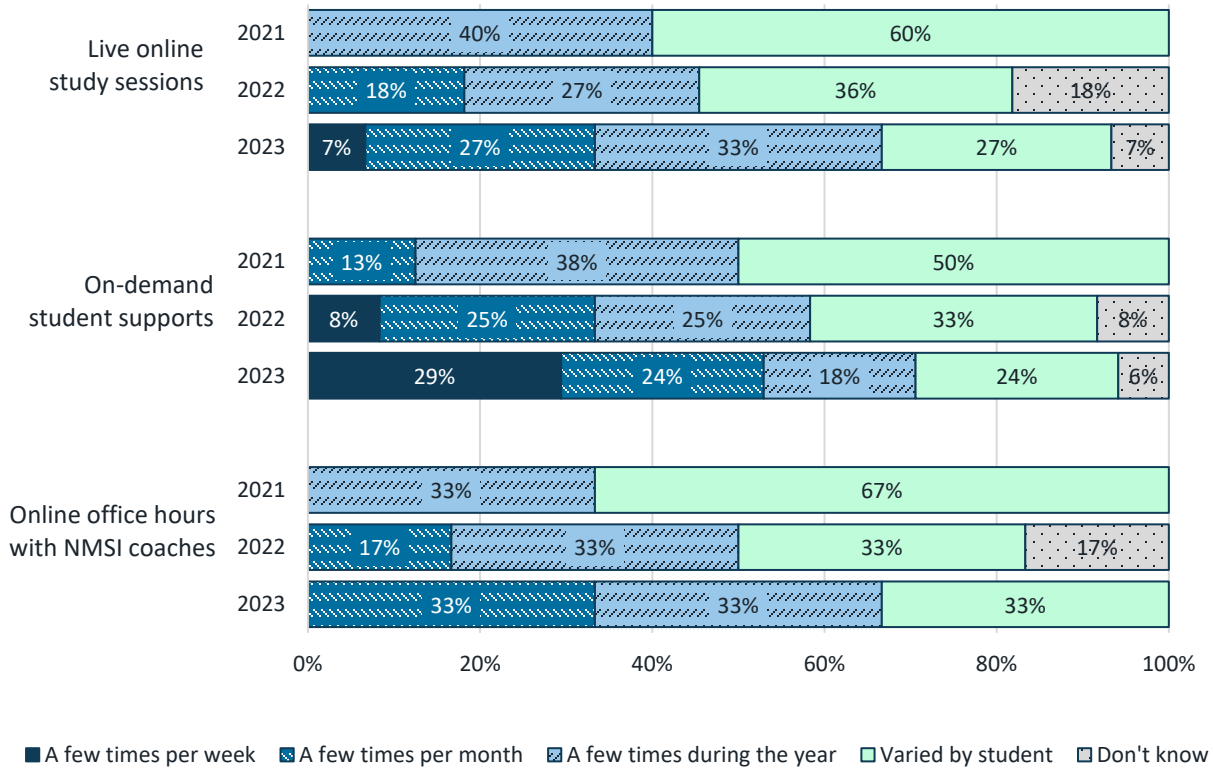
Finally, in Year 3, 18 teachers (60%) reported student use of at least one of the resources. Half of the teachers (50%) reported that at least a few students had used the live online student sessions (bottom bar of top group in Exhibit 3.4), and more than half (59%) reported that at least a few students had used the on-demand student supports. Student participation in the online office hours remained generally low in Year 3 (with 31% of teachers reporting any students) but was higher than in either previous year (21% in Year 1 and 26% in Year 2). And six teachers—up from three the previous year and one in Year 1—reported that most or all students had used one or more of the three resources.

Among the 12 teachers for whom we have data for all three years, only three reported any use of the student resources in Year 1, while six reported use in Year 2 and seven reported use in Year 3. With just a single exception (one teacher from Year 1 to Year 2), every teacher who reported use of the student resources in a given year continued to report such use in the subsequent year. Therefore, the apparent increase over time in use of the student resources is likely not solely a function of who responded to the survey. That is, some teachers who did not have their students use the resources early on did start to use them later.

Teachers who reported that any students used a resource were asked how often students used that resource. As with the numbers of teachers and students, the frequency of use also increased over time. The most common response in Year 1, selected by at least half of the teachers for each of the three resources, was that it varied by student (top bar of each group in Exhibit 3.5). No teacher in Year 1 indicated that any of the three resources was used more often than a few times per month. In fact, only a single teacher selected "a few times per month"—and only for the on-demand student supports. The other teachers reported student use of each resource only a few times per year.

In Year 2 (middle bar of each group in Exhibit 3.5), only about one third of teachers reported that frequency of use varied by student, compared with one half to two thirds in Year 1. Although teachers were more likely in Year 2 to report that they did not know how often students used the resources, those who did know were more likely to report that use occurred more frequently, compared with Year 1. In particular, while only one of the resources was reported to be used as often as a few times per month in Year 1 (and only one teacher reported that frequency of use for that resource), all three resources were reported by some teachers to be used a few times per month in Year 2 (with more than one teacher reporting that frequency of use for two of the three resources). In fact, in Year 2, there was even one teacher who reported use of one of the resources (on-demand supports) as often as a few times per *week*.

Exhibit 3.5. Teacher Report on How Often Students Used the Student Resources



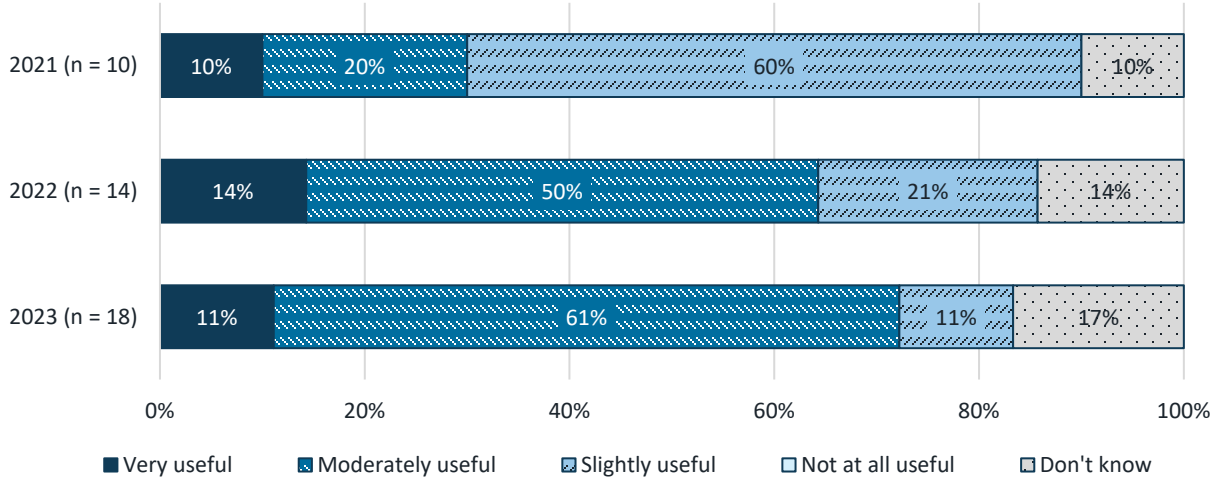
Note. In Year 1 (2021), *n*s for the three bars are 10, 8, and 6 respectively (from top to bottom). In Year 2 (2022), *n*s for the three bars are 11, 12, and 6 respectively (from top to bottom). In Year 3 (2023), *n*s for the three bars are 15, 17, and 9 respectively (from top to bottom).

Frequency of use continued to increase in Year 3 (bottom bar of each group in Exhibit 3.5). Perhaps most notably, five teachers (29%) reported that students used the on-demand supports a few times per week, up from one teacher in Year 2 (8%) and none in Year 1. For each of the three resources, large majorities of the responding Year 3 teachers reported use at least a few times per year: 67% for live online study sessions (compared to 40% in Year 1 and 45% in Year 2), 71% for on-demand student supports (compared to 50% in Year 1 and 58% in Year 2), and 67% for online office hours (compared to 33% in Year 1 and 50% in Year 2). Correspondingly, Year 3 teachers were less likely than Year 1 or 2 teachers to report that frequency of use varied by student or that they did not know. This might suggest that by Year 3, teachers were being more intentional about having their students use the resources (e.g., by assigning use of the resources rather than allowing use to be optional).

Perceptions of Usefulness of the Student Resources

Teachers who indicated that at least one resource was used by at least a few students were asked how useful students had found the content of the resources to be. As shown in Exhibit 3.6, perceptions of usefulness increased from year to year. In Year 1, only 30% of teachers (three of 10) reported that students had found the student resources to be at least moderately useful; the corresponding percentages in Years 2 and 3, respectively, were 64% (nine of 14 teachers) and 72% (13 of 18 teachers). In none of the three years did any teachers report that students had found the resources to be not at all useful, and the percentage reporting only slightly useful dropped dramatically from 60% in Year 1 to 21% in Year 2 to 11% in Year 3. It stands to reason that increased use of the student resources, as noted in the previous subsection, would be accompanied by increased perceptions of the resources' usefulness.

Exhibit 3.6. Teacher Report on How Useful Students Found the Student Resources



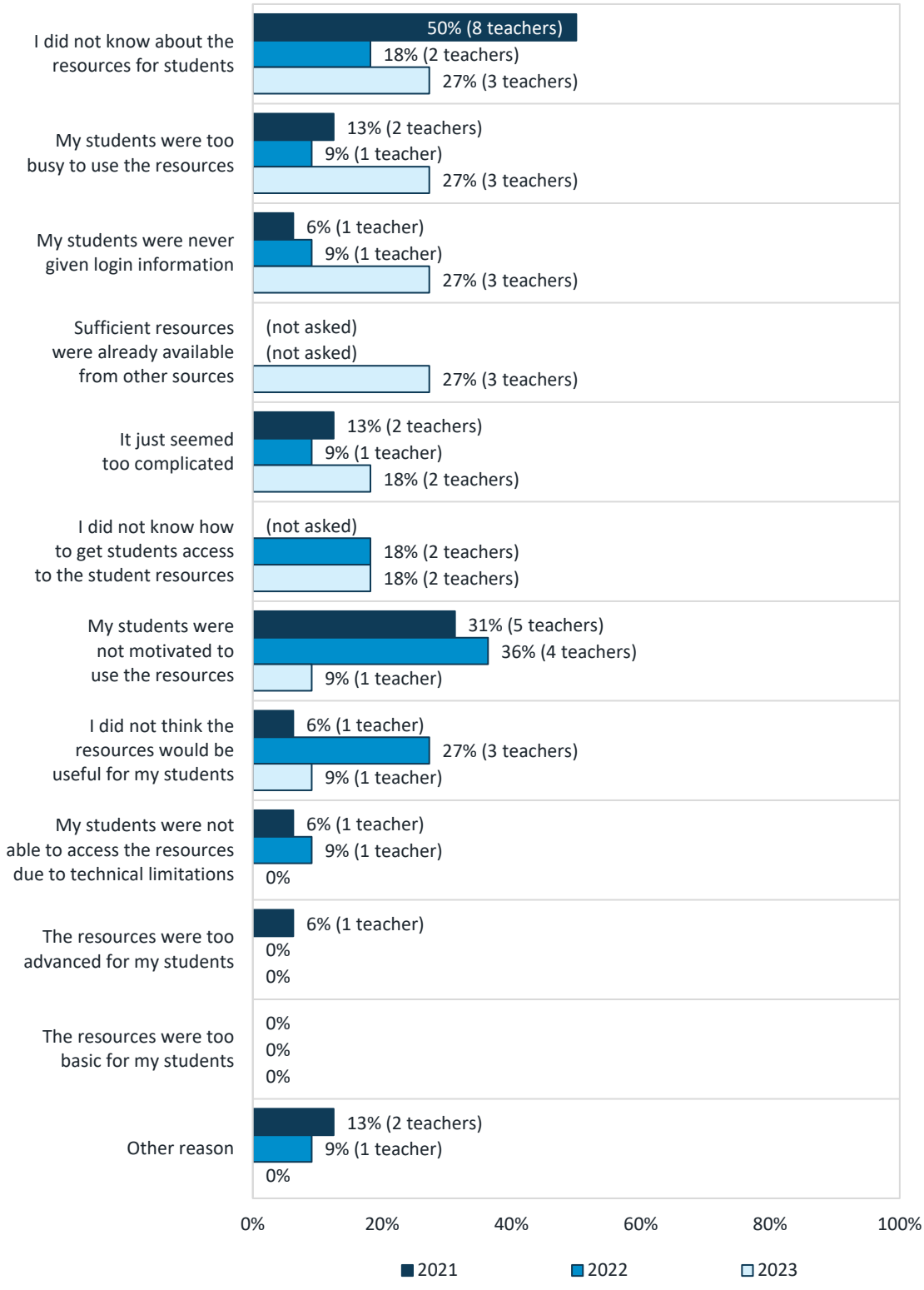
Reasons for Low Use of the Student Resources

Teachers who reported little or no use of the student resources—defined as no more than a few students and no more than a few times per year for any of the three resources—were asked why use of the student resources was not greater. In Year 1, as with the activities for teachers themselves, the most common reason given for low use of the student resources in Year 1 was that the teachers *did not know about these resources*—selected by fully half of the 16 teachers who indicated low use (Exhibit 3.7). The next most common reason, selected by about one third of the teachers, was that students were not motivated to use the resources.

We see a somewhat different pattern in Year 2, when only 11 teachers indicated low use of the student resources. Unlike the 50% of teachers in Year 1 who reported that lack of use of the student resources resulted from not knowing about the resources, only 18% of teachers in Year 2 marked this as a reason. In Year 2, the most commonly reported reason was that students were not motivated to use the resources (marked by 36% of the teachers). Not far behind was that the teacher did not think the resources would be useful for their students (marked by 27% of teachers, although that translates to only three teachers, given the small *N*).

The pattern shifted again somewhat in Year 3. Among the 11 teachers who indicated low use of the student resources, there was a four-way tie for the most commonly selected reason: three teachers each (27%) indicated that they did not know about the resources, that their students were too busy, that the students were never given login information, and (in a new option not offered in Year 1 or 2) that sufficient resources for students were already available from other sources. Compared to Years 1 and 2, many fewer teachers in Year 3 (in fact, only one) said that the students were not motivated to use the resources; similarly, only one teacher in Year 3 said that they did not think the resources would be useful.

Exhibit 3.7. Teacher Report of Reasons Why Use of the Student Resources Was Low

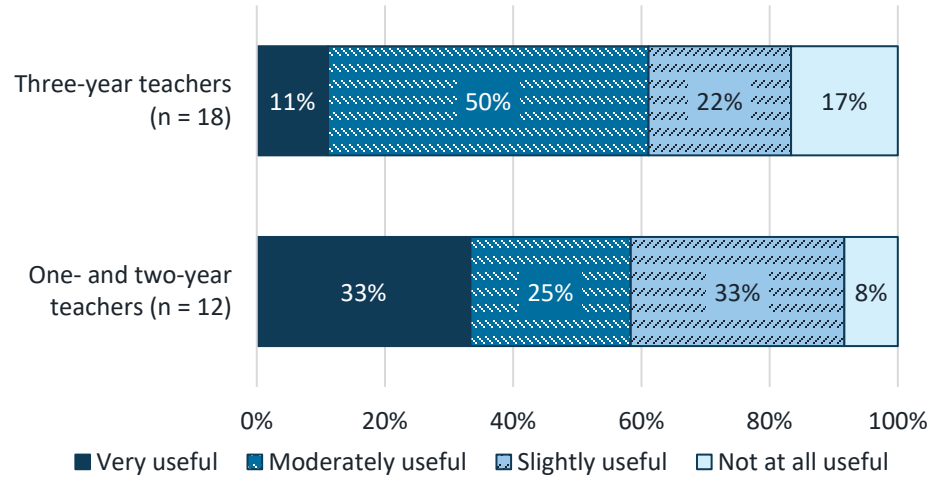


Note. In Year 1, *n* = 16. In Years 2 and 3, *n* = 11.

Teacher Perceptions of the Lone Star AP CSP Program as a Whole

The Year 3 survey asked the Lone Star AP CSP respondents how useful, overall, they had found the program.[8] All 30 teachers answered the question, with 20% saying *very useful*, 40% saying *moderately useful*, 27% saying *slightly useful*, and 13% saying *not at all useful*. Respondents who had been in the program all three years were most likely to select moderately useful (50%), while for teachers who joined in Year 2 or Year 3, responses were relatively evenly split among the first three categories (Exhibit 3.8).

Exhibit 3.8. Teacher Opinion About Overall Usefulness of Lone Star AP CSP Program



Teachers' Reflections on the Three Years of the Program

As noted earlier in this chapter, the Year 3 survey asked the 18 teachers who were in all three years' program samples an additional set of questions about their impressions of and experiences with the program over the course of its three years. The 18 teachers were fairly evenly split in how they described their overall level of participation in the program across the three years:

- Three (17%) marked high in all three years
- Three (17%) marked medium in all three years
- Three (17%) marked low in all three years
- Four (22%) marked higher at first, then lower
- Two (11%) marked lower at first, then higher
- Three (17%) varied from year to year without a clear pattern

[8] The Year 1 and Year 2 surveys instead asked teachers how likely they were to continue in the program and how enthusiastic they were about continuing. Responses to these questions were summarized in Jones and Holtzman (2023).

This high degree of variation is interesting in and of itself and also provides important context for interpretation of responses to the subsequent questions, which are described next and illustrated in Exhibit 3.9. In particular, teachers with lower levels of participation may not have accurate perceptions of the program's offerings. On the other hand, it is also possible that teachers with lower levels of participation were unsatisfied with the available offerings, thereby explaining their lower participation.

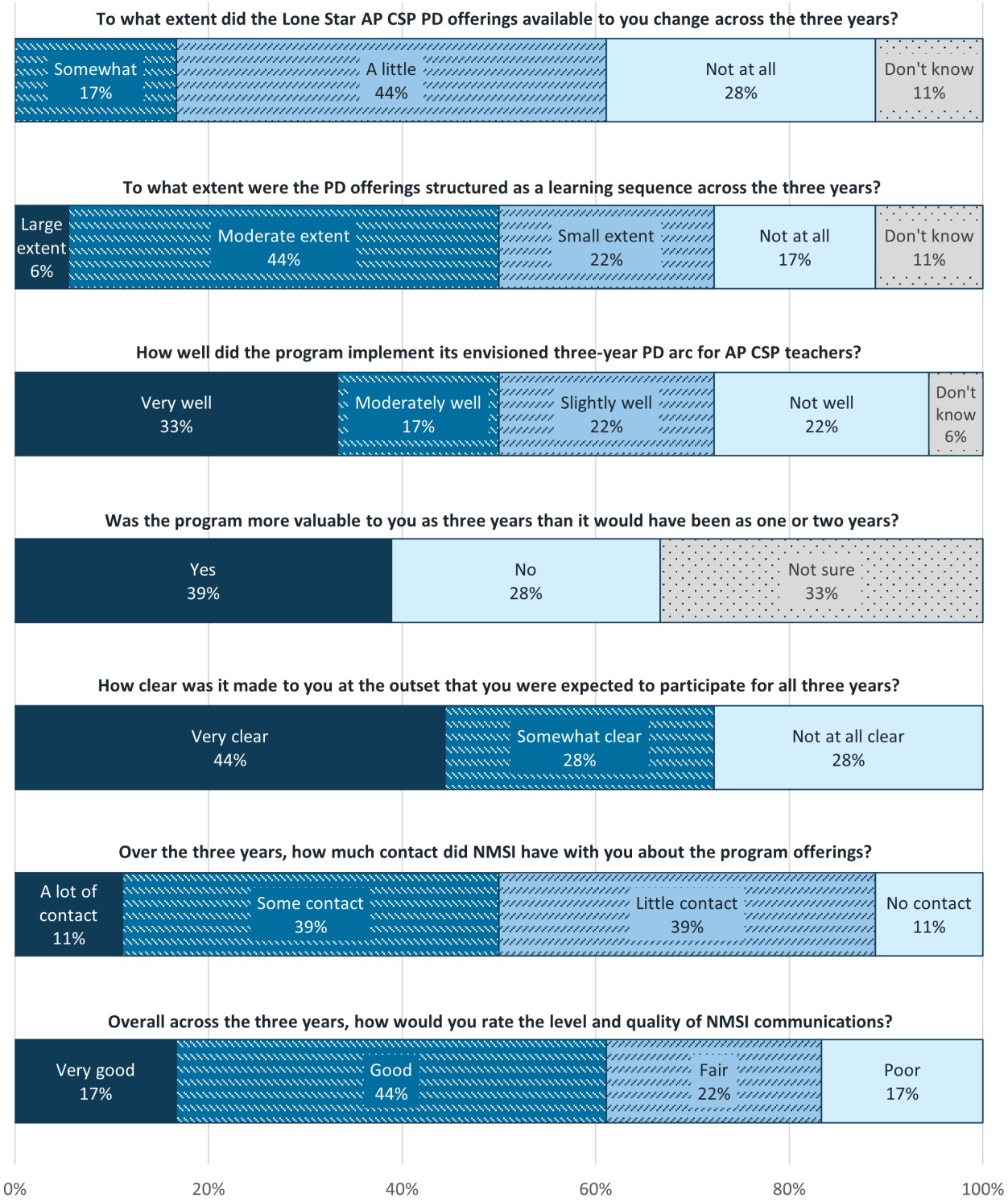
The next survey question was "To what extent did the Lone Star AP CSP professional development offerings (courses, etc.) available to you change across the three years?" None of the 18 teachers indicated that the offerings had changed a lot; three (17%) said somewhat, eight (44%) said a little, five (28%) said not at all, and two (11%) marked *don't know*. To the extent that the program was meant as a three-year arc of professional development, with each year building on the previous year, one might have expected somewhat greater perception of change across the three years. In fact, the following question asked about this more directly:

> *The Lone Star AP CSP program was originally envisioned as a three-year training arc, sequence, or cycle for teachers, involving "a sustained teacher professional development [PD] cycle which responds to teachers' learning progressions over time." To what extent were the Lone Star AP CSP professional development offerings structured as a* learning sequence *across the three years?*[9]

The most common response, given by eight teachers (44%), was *to a moderate extent*. Four teachers (22%) marked *small extent* and three teachers (17%) marked *not at all*. Only one teacher (6%) marked *large extent*, and two teachers marked *don't know* (11%). Thus, there appears to have been quite a range in perception. The subsequent question, "How well do you think the program implemented its envisioned *three-year professional development arc* for AP CSP teachers?" was even more direct, and also yielded a range of responses. The modal response was *very well* (given by six teachers, 33%), but four teachers (22%) marked *not well*, and another four (22%) marked *slightly well*. Three (17%) said *moderately well*, and one (6%) said *don't know*.

[9] The quoted portion of the survey question is from the program's EIR proposal (NMSI, 2019).

Exhibit 3.9. Three-Year Teachers' Opinions on the Three-Year Nature of the Program



Note. $N = 18$ for all questions.

Then, asked "As a three year program, do you think the Lone Star AP CSP program was more valuable to you, as a teacher, than it would have been if it had only been one or two years?", teachers' responses were relatively evenly split: seven (39%) said yes, five (28%) said no, and six (33%) said they weren't sure. Responses to the next question, "How clear was it made to you at the outset of the program (summer or early fall 2020) that you were expected to participate for all three years of the program?", were also fairly mixed, with eight (44%) saying very clear, five (28%) saying somewhat clear, and another five (28%) saying not at all clear.

This question served as a segue to the survey section's final two questions, which were about communication from NMSI personnel. The first of these, "Over the course of the three years, how much contact did NMSI personnel have with you about Lone Star AP CSP offerings?", had a more typical distribution of responses, with two teachers (11%) saying no contact, seven (39%) saying little contact, seven (39%) saying some contact, and two (11%) saying a lot of contact. Finally, the teachers were asked to rate the level and quality of communication from NMSI program personnel, overall across the three years, about the Lone Star AP CSP program offerings. Again, responses were mixed: three (17%) said very good, eight (44%) said good, four (22%) said fair, and three (17%) said poor.

In sum, there was a distinct lack of consistency among the 18 three-year teachers in their perceptions of the program—responses to nearly all of the questions could well be described as "all over the map." It is thus somewhat difficult to know what to make of the data. Ideally, we would conduct additional analyses to look for associations between the perceptions and factors such as level of program participation or teaching experience, but with only 18 teachers total, it would be difficult to detect systematic differences.

Summary and Discussion

In summary, Lone Star AP CSP teachers' responses on surveys administered at the end of each of the three years of the program suggest the following:

- Teacher participation in Lone Star AP CSP activities appears to have increased from Year 1 to Year 2. There was not much change from Year 2 to Year 3 in participation levels, but perceptions of usefulness of the teacher activities increased across the three years.
- Use of the student resources, as reported by teachers, increased from year to year, as did teacher perceptions of the usefulness of these resources.
- Lack of awareness about the program's supports and resources was a commonly cited reason for nonuse of particular activities in all three years, but more so in Year 1 than Year 2, and more so in Year 2 than Year 3.

- At the close of the program in Year 3, most teachers (60%) reported that overall, they had found the program to be at least moderately useful.
- The subset of teachers who participated in all three years of the program had mixed opinions about the cohesiveness of the program across the three years, such as how well it implemented the envisioned three-year arc of professional development.

A serious caveat of all of these findings, however, is that they are based on teachers' self-report on surveys, and the self-report may have been inaccurate. For instance, teachers may have misremembered what activities they had participated in—particularly from the previous summer, which was nearly a year prior to their survey completion. Perhaps even more importantly, not all Lone Star teachers responded to the surveys, and those who did may not have been representative of the full sample. In particular, it is plausible that the teachers who responded to the survey may have been more engaged with the program than those who did not. If this is the case, the survey results may overstate teachers' levels of use of the program, as well as their perceptions of its usefulness, in the program sample as a whole.

Perhaps partly for this reason, the uptake findings seem somewhat more optimistic than the fidelity-of-implementation findings reported in the previous chapter, particularly for the teacher supports component in Years 2 and 3. Unlike the uptake analysis, which was based on a self-selected sample of teachers (those who chose to respond to the survey), the fidelity analysis covered the entire program sample. Other factors that might explain the differences between the two sets of findings are as follows:

- The uptake analysis looked at whether teachers participated in particular program elements at all, whereas the fidelity analysis examined whether they participated at a particular level (e.g., earned a minimum number of credits).
- The elements included in the two analyses differed to some extent. For example, the uptake analysis separated out the different elements of the summer training, and also looked at some elements, such as the Teacher Villages, that were not included in the fidelity analysis.
- Because the uptake analysis was based on teacher survey data, it only examined the supports for teachers and (indirectly) the supports for students components, and not the supports for schools component, which the fidelity analysis found was not implemented with fidelity in the second and third years.
- For the student supports, the fidelity analysis examined only whether the supports were available (a relatively low bar), whereas the uptake analysis attempted to gauge the extent to which the student resources were actually used (albeit as reported by teachers).

Moreover, even though the uptake analysis found increased participation in some of the teacher PD elements of the program in Years 2 and 3, this is due in part to the low levels of participation in Year 1. Even in the second and third years, participation levels did not exceed 80% of responding teachers for any of the elements, and responding teachers constituted only 57% of the Year 2 sample and 63% of the Year 3 sample. Participation levels likely were lower among the nonrespondents, although we do not know for sure.

In addition, teacher lack of awareness of the program and its elements was an issue in all three years of the program, though admittedly less so (at least, according to responding teachers) as time went on. One would have hoped that as schools increasingly returned to "normal" following the greatly pandemic-disrupted 2019–20 and 2020–21 school years, NMSI program staff might have made greater efforts to personally reach out to the Lone Star AP CSP teachers—as well as to school leaders and counselors, with respect to some of the fidelity findings reported in the previous chapter—and encourage them to take advantage of the program's resources.

Only with relatively high levels of engaged participation would we reasonably expect the Lone Star AP CSP program to have its theorized impacts, both on intermediate outcomes such as teachers' instructional practice and on the more distal outcomes of student participation in and performance on the AP CSP exam. These are the topics of the remaining chapters of this report.

Chapter 4. Impact on Instructional Practices

As detailed in previous chapters, instructional practice in the AP CSP course was one of the key mediators by which the program aimed to have an impact on students. Accordingly, AIR's evaluation studied the impact of the program on teachers' instructional practices; this chapter describes that study and its findings. The research question addressed is as follows:

3. Did participation in the Lone Star AP CSP program improve AP CSP teachers' instructional practices in the AP CSP course, compared to AP CSP teachers in the business-as-usual condition, after one year, after two years, and after three years?

The study used a quasi-experimental matched comparison group design and an analysis-of-covariance (regression) estimation model to examine the effect of Lone Star AP CSP on participating teachers' instructional practices, as measured by teachers' self-report on surveys.

The remainder of this chapter is organized as follows. First, we describe the measures of instructional practice that formed the basis for our sample selection and subsequent analysis. We then explain how the teacher samples were selected in each analysis year, including how we matched Lone Star teachers to comparison teachers and the extent to which the two groups were similar at baseline. We then detail our statistical estimation strategy, followed by presentation of the results. The chapter concludes with a brief summary and discussion, including acknowledgment of the limitations of the teacher impact study.

KEY FINDINGS

- After one year of implementation, the Lone Star AP CSP program had a significant positive effect on the extent to which teachers encouraged students to use cognitive problem-solving strategies.
- After two years of implementation, the program had a significant and large positive effect on each of the three measures of instructional practice we examined: Engaging Students in Practices of Computer Science, Encouragement of Cognitive Problem-Solving Strategies, and Use of Pedagogically Based Learning Strategies.
- After three years of implementation, the program had no effect on any of the three measures of instructional practice.

Instructional Practice Measures

The measures of instructional practice examined for this study come from surveys of AP CSP teachers administered by AIR and NMSI in summer 2020, spring 2021, spring 2022, and spring

2023.[1] (The following section, "Teacher Sample," includes information about how the surveyed teachers were selected.) Several survey questions asked the respondents about the instructional practices they had been using in their AP CSP course over the course of the school year.

After the AIR survey was administered for the first time in summer 2020, we conducted factor analysis and Rasch modeling on the computer science instructional practice item responses to identify constructs that not only made sense conceptually but also were psychometrically reliable.[2] Based on these analyses, we formulated three separate measures to serve as outcomes for the study of impact on instructional practices:

- Engaging Students in Practices of Computer Science
- Encouragement of Cognitive Problem-Solving Strategies
- Use of Pedagogically Based Learning Strategies

Each of the three measures consists of multiple items, all answered on a frequency scale; see the box in this section for the individual items constituting each measure. The items in the first measure were taken from the High School Computer Science Teacher Questionnaire used in the 2018 National Survey of Science and Mathematics Education (NSSME; Banilower et al., 2018) funded by the National Science Foundation (NSF). This set of items was designed to gauge respondents' use of instructional practices related to high school computer science generally. The items constituting the second and third measures were developed by the AIR evaluation team based closely on the instructional strategies described in the *AP CSP Course and Exam Description* in effect starting in the 2020–21 school year (College Board, 2020) and thus were tailored for AP CSP specifically.

[1] AIR developed the survey instrument, managed the online platform (including the collection of response data from both Lone Star and potential comparison teachers), and sent invitations and reminders to potential comparison teachers. NMSI sent invitations and reminders to the Lone Star teachers, because NMSI had promised these teachers that their contact information would not be shared beyond NMSI. However, AIR provided the text for these communications and asked NMSI to send them on the same schedule as the communications for the potential comparison teachers. Therefore, the data collection procedures were consistent. The 2021, 2022, and 2023 surveys also included questions for the Lone Star teachers about their participation in and perceptions of program activities (see Chapter 3); potential comparison teachers were routed out of these sections.
[2] A total of 103 teachers completed the instructional practice items in 2020. See the "Teacher Sample" section of this chapter for more information about the sample.

ITEMS CONSTITUTING THE INSTRUCTIONAL PRACTICE MEASURES

Measure 1: Engaging Students in Practices of Computer Science

How often did you have students do each of the following in your [school year] AP CSP class(es)? *(Never, A few times during the year, Once or twice a month, Once or twice a week, All or almost all lessons)*

- Create computational artifacts (for example: programs, simulations, visualizations, digital animations, robotic systems, or apps)
- Create a computational artifact designed to be used by someone outside the class or other students
- Provide feedback on other students' computational products or designs
- Get input on computational products or designs from people with different perspectives (do not include feedback that you give students)
- Systematically use test cases to verify program performance and/or identify problems
- Identify real-world problems that might be solved computationally
- Consider how a program they are creating can be separated into modules/procedures/objects
- Identify and adapt existing code to solve a new computational problem
- Use computational methods to simulate events or processes (for example: rolling dice, supply and demand)
- Analyze datasets using a computer to detect patterns
- Write comments within code to document purposes or features
- Create instructions for an end-user explaining how to use a computational artifact
- Explain computational solution strategies verbally or in writing
- Compare and contrast the strengths and limitations of different representations such as flow charts, tables, code, or pictures

Source: NSSME High School Computer Science Teacher Questionnaire (Banilower et al., 2018)

Measure 2: Encouragement of Cognitive Problem-Solving Strategies

How often did you have students do each of the following in your [school year] AP CSP class(es)? *(Never, A few times during the year, Once or twice a month, Once or twice a week, All or almost all lessons)*

- Go step-by-step through program code by hand to determine how a piece of program code operates
- Break a problem into smaller pieces, parts, or subtasks
- Create a plan or develop a process for solving a problem
- Identify, troubleshoot, and correct errors
- Identify patterns or trends and draw conclusions from them
- Make predictions or conjectures about what outcomes an activity will have, and then confirming or modifying the conjectures based on observed outcomes
- Make an abstract problem more concrete by using simpler cases or examples
- "Think aloud"/talk through a difficult problem (what the problem is and how it might be solved)
- Work backward from an ideal solution to help solve a problem
- Recall what they already know about a concept before extending the concept or taking it in a new direction

Source: Adapted by AIR from the AP CSP Course and Exam Description (College Board, 2020)

Measure 3: Use of Pedagogically Based Learning Strategies

How often did you have students do each of the following in your [school year] AP CSP class(es)? *(Never, A few times during the year, Once or twice a month, Once or twice a week, All or almost all lessons)*

- Highlight, underline, or annotate text to focus on key information and help understand the text or solve the problem
- Work in pairs to write and review/revise program code, switching roles every so often
- Engage in interactive, small-group discussions with one another to hear new perspectives that might provide insights about how to solve a problem
- "Jigsaw," where students form groups to learn about particular topics and then form new groups composed of one "expert" from each topic to teach the others
- Use body movements (e.g., standing and sitting) or hands-on materials to convey or foster understanding of a concept
- Discuss their work with one another, including providing feedback to one another
- Start off thinking through a problem alone, then pair with a partner to share and refine each other's ideas, and then share the refined ideas with the whole class
- Use non-electronic games, puzzles, or manipulatives to provide new insights into a concept and foster understanding of it
- Create a visual representation (e.g., diagram or flowchart) to organize information or create a plan for solving a problem
- Keep a journal or log of their progress on a particular project, including accomplishments, challenges, and potential next steps
- Write for a short, pre-specified amount of time (e.g., 5 minutes) in response to a specific prompt
- Record, maintain, and reinforce vocabulary words using a graphic organizer including definitions, pictures, notations, or connections

Source: Adapted by AIR from the AP CSP Course and Exam Description (College Board, 2020)

Because the survey items had ordinal response categories that were not necessarily equidistant between points, AIR used a Rasch model for ordered response categories (Andrich, 1978; Rasch, 1980; Wright & Masters, 1982) to scale the survey items constituting each of the three measures. The scaling sample was the full set of 103 teachers who had completed the instructional practice items on the 2020 survey (see next section, "Teacher Sample"). The logit scores produced by the Rasch modeling were used in the analysis of the impact of Lone Star AP CSP on teachers' instructional practices.[3]

As shown in Exhibit 4.1, all three measures have high reliability, indicating that teachers tended to answer the items in each measure in internally consistent ways. Correlations among the three measures are provided in Appendix Exhibit C-1.

[3] In the baseline year, the logit scores ranged from -3.92 to 6.54 for Measure 1, from -4.22 to 6.43 for Measure 2, and from -4.18 to 5.45 for Measure 3. The Rasch-Andrich thresholds from the Rasch models in the baseline year were used to calibrate the scores in Rasch models in each of the following three outcome years. This "anchoring" of survey responses from the outcome years on data from the baseline year was done to make the scores comparable across years.

Exhibit 4.1. Reliability of the Instructional Practice Measures

Measure	Number of items	Rasch/person reliability	Cronbach's alpha
Measure 1: Engaging Students in Practices of Computer Science	14	0.90	0.91
Measure 2: Encouragement of Cognitive Problem-Solving Strategies	10	0.86	0.91
Measure 3: Use of Pedagogically Based Learning Strategies	12	0.86	0.87

Note. Reliabilities were calculated using the 2020 scaling sample. *N* = 102 for the first measure and 103 for the second and third measures.

Teacher Sample

The three analyses described in this report—impact after one year, two years, and three years of implementation—differed somewhat in which teachers were included in the sample for each. All three drew on a common pool of teachers who had completed the baseline (2020) survey, but they diverged based on which of these teachers completed the 2021, 2022, and 2023 surveys, given that the surveys in these three years provided the outcome data.

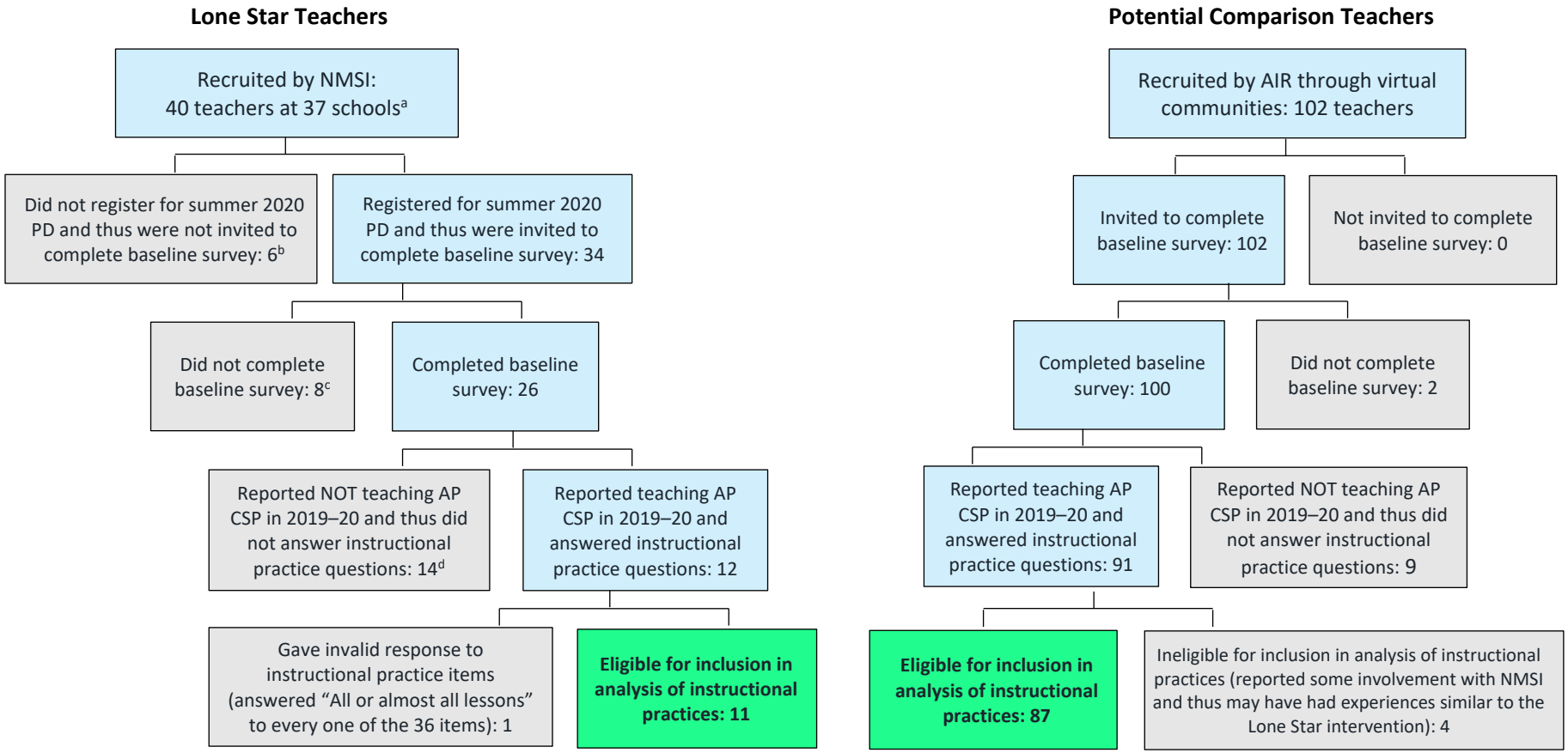
Baseline Year (Full Pool of Eligible Teachers)

The 2020 survey preceded the initial professional development NMSI provided to Lone Star AP CSP teachers, so the measures from this 2020 survey administration serve as baseline (pre-intervention) measures of the instructional practice outcomes. The survey was administered online to the 34 Lone Star teachers who had registered for NMSI's summer training, along with 102 potential comparison teachers recruited by AIR from virtual communities of computer science teachers.[4]

To be eligible for inclusion in the study of instructional practices, survey respondents had to have taught AP CSP during the 2019–20 school year and to have provided valid responses to the questions about their instructional practices in the course that year. This criterion was met by 11 Lone Star teachers (from 10 schools) and 87 potential comparison teachers. Exhibit 4.2 provides further detail on how these numbers were derived.

[4] There were two such communities: the AP CSP Open group of *CS for All Teachers* (an NSF-funded and AIR-managed project for computer science educators) and a Facebook group for AP CSP teachers. We posted announcements and teachers had to complete an eligibility screening questionnaire before they were invited to complete the actual survey. The selected teachers were located throughout the United States. An analysis of school names and street addresses, as teachers filled them in on the screening instrument, indicated that only two of the potential comparison teachers taught at the same school.

Exhibit 4.2. Baseline Pool Consort Diagram



[a] Initially we thought there were 41 Lone Star teachers, but we later learned that one of the 41 was in a non-teaching role at their school.

[b] We learned in a later year's survey that two of these six teachers *did* teach AP CSP in 2019–20. (We do not know about the other four.) Unfortunately, without the baseline survey data, we could not include them in the analysis of instructional practices.

[c] We learned in a later year's survey that two of these eight teachers *did* teach AP CSP in 2019–20. Unfortunately, without the baseline survey data, we could not include them in the analysis of instructional practices. (Of the other six, we do not know about three of them, and three indicated that they had not taught AP CSP in 2019–20 or in any earlier year.)

[d] In fact, all 14 of these teachers indicated that they had *never* taught the AP CSP course.

Sample for Analysis of Impact After One Year of Implementation (2021)

The 2020–21 school year was the first year of Lone Star implementation. The spring 2021 survey included the same set of instructional practice items that had been on the 2020 survey, except that they asked the teachers about their AP CSP instruction during the 2020–21 year rather than the 2019–20 year. The 2021 instructional practice items were completed by nine of the 11 eligible Lone Star teachers and 77 of the 87 potential comparison teachers. All nine of the Lone Star teachers, who taught at nine different schools, were included in the Year 1 analysis of instructional practices. From the 77 potential comparison teachers, we used a matching process (see next section) to select 54 of them (i.e., a 1:6 matching ratio[5]) for inclusion in the analysis. Three of the selected comparison teachers were in Texas, so state was not confounded with the intervention (given that all of the Lone Star teachers were in Texas).

Sample for Analysis of Impact After Two Years of Implementation (2022)

The 2021–22 school year was the second year of Lone Star implementation. The instructional practice items (now referencing the 2021–22 school year) were included again in the 2022 survey. The items were completed by seven of the 11 eligible Lone Star teachers and 66 of the 87 potential comparison teachers. All seven of the Lone Star teachers (who again each taught at a different school) were included in the Year 2 analysis of instructional practices, and from the 66 potential comparison teachers, we matched at a 1:7 ratio to select 49 of them for inclusion in the analysis. As in the prior year's sample, three of the selected comparison teachers were in Texas.

Sample for Analysis of Impact After Three Years of Implementation (2023)

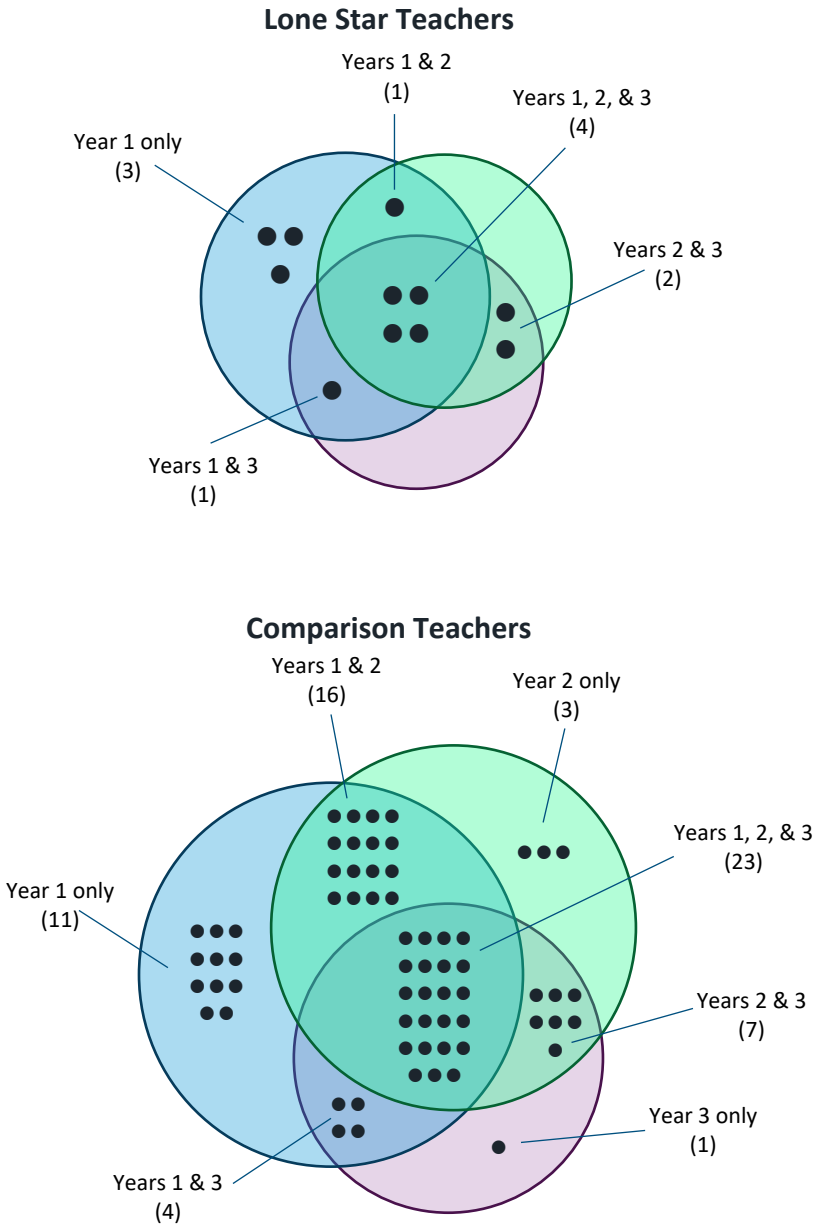
The 2022–23 school year was the third and final year of Lone Star implementation. The instructional practice items (now referencing the 2022–23 school year) were included yet again. The items were completed by seven of the 11 eligible Lone Star teachers and 55 of the 87 potential comparison teachers. All seven of the Lone Star teachers, who were from six different schools, were included in the Year 3 analysis of instructional practices, and from the 55 potential comparison teachers, we matched at a 1:5 ratio to select 35 of them for inclusion in the analysis. One of them was in Texas.

Degree of Overlap Among the Three Samples

It may be worth noting that the samples for the Year 2 analysis and the Year 3 analysis, though smaller than the Year 1 analysis sample, were not subsets of it. In addition, the seven Lone Star teachers in the Year 2 sample were not the same seven teachers in the Year 3 sample. Exhibit 4.3 shows the extent of overlap among the three samples.

[5] In each year, we used the highest matching ratio we could find that would yield baseline equivalence.

Exhibit 4.3. Overlap Among the Three Years' Analysis Samples



Matching Process and Baseline Equivalence

To select the comparison teachers to be included in each year's analysis—in particular, a subset of the potential comparison teachers who closely resembled the Lone Star teachers at baseline—we used a matching technique called *scaled Euclidean distance matching* (Judkins, 2013). This procedure allows for differential weighting of different matching variables—in our case, the three instructional practice measures. We employed a procedure of iteratively testing different weight sets to optimize baseline equivalence on each of the three measures. We also explored different ratios of Lone Star teachers to comparison teachers, with the goal of having

as many teachers in the analysis as possible, for statistical power purposes, while meeting baseline equivalence criteria.

Our measure of baseline equivalence was Hedges' *g,* including a correction (ω) for small sample sizes. This measure calculates a *standardized mean difference* (SMD) between groups, i.e., a difference expressed in terms of standard deviation units (also called an effect size). See the box for the formula and the meaning of its terms in our baseline equivalence calculations.

FORMULA FOR HEDGES' *G* AT BASELINE

The formula for Hedges' *g* is as follows:

$$\text{Hedges' } g = \frac{(X_i - X_c)}{\sqrt{\frac{(n_i - 1)SD_i^2 + (n_c - 1)SD_c^2)}{n_i + n_c - 2}}} \times \omega$$

At baseline, the numerator of this formula is the difference between the group means, i.e.:

X_i is the (unadjusted) Lone Star group mean on the outcome measure at baseline (2020);

X_c is the (unadjusted) comparison group mean on the outcome measure at baseline (2020).

The denominator is the pooled standard deviation, where:

SD_i is the Lone Star group standard deviation of the outcome measure at baseline (2020);

SD_c is the comparison group standard deviation of the outcome measure at baseline (2020);

n_i is the number of Lone Star teachers in the analysis; and

n_c is the number of comparison teachers in the analysis.

Finally, ω is the small sample size correction, calculated as:

$$\omega = 1 - \frac{3}{4N - 9}$$

where $N = n_i + n_c$ (i.e., total sample size).

Baseline Equivalence for Analysis of Impact After One Year of Implementation

For the sample of comparison teachers to be included in the analysis of impact after one year of program implementation, a 1:6 matching ratio yielded a set of comparison teachers who, as a group, were equivalent to the Lone Star teachers at baseline on all three instructional practice measures. As shown in Exhibit 4.4, the baseline SMDs were all less than (an absolute value of) 0.25 standard deviations. Given that the statistical models for the impact analysis include the baseline measures as covariates (see the "Analysis Method" section), baseline SMDs of this magnitude satisfy criteria set by the What Works Clearinghouse (2022) for baseline equivalence.

Exhibit 4.4. Baseline Statistics for the Year 1 Analysis Sample

Measure	Lone Star teachers (*n* = 9)	Comparison teachers (*n* = 54)	*p* value	Hedges' *g*
Engaging Students in Practices of Computer Science	-0.373 (1.465)	-0.514 (1.116)	0.739	0.119
Encouragement of Cognitive Problem-Solving Strategies	0.599 (2.300)	0.784 (1.455)	0.748	-0.115
Use of Pedagogically Based Learning Strategies	-0.370 (1.703)	-0.171 (0.919)	0.601	-0.187

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from independent-samples *t* tests; they are provided for parallelism with the outcome analysis results.

Baseline Equivalence for Analysis of Impact After Two Years of Implementation

For the analysis of impact after two years of program implementation, we were able to achieve baseline equivalence with a 1:7 matching ratio, yielding 49 comparison teachers, helping to offset the smaller number of Lone Star teachers (seven) in this analysis. Again, the baseline SMDs were below (an absolute value of) 0.25 for all three of the instructional practice measures (Exhibit 4.5).

Exhibit 4.5. Baseline Statistics for the Year 2 Analysis Sample

Measure	Lone Star teachers (*n* = 7)	Comparison teachers (*n* = 49)	*p* value	Hedges' *g*
Engaging Students in Practices of Computer Science	-0.097 (1.872)	-0.236 (1.268)	0.799	0.102
Encouragement of Cognitive Problem-Solving Strategies	0.623 (2.409)	0.733 (1.399)	0.860	-0.070
Use of Pedagogically Based Learning Strategies	-0.203 (2.074)	-0.067 (0.936)	0.765	-0.120

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from independent-samples *t* tests; they are provided for parallelism with the outcome analysis results.

Baseline Equivalence for Analysis of Impact After Three Years of Implementation

For the analysis of impact after three years of program implementation, we had to use a lower matching ratio—1:5, with seven Lone Star teachers and 35 comparison teachers—to achieve baseline equivalence. As shown in Exhibit 4.6, all three baseline SMDs were still below the 0.25 threshold.

Exhibit 4.6. Baseline Statistics for the Year 3 Analysis Sample

Measure	Lone Star teachers (*n* = 7)	Comparison teachers (*n* = 35)	*p* value	Hedges' *g*
Engaging Students in Practices of Computer Science	-0.283 (1.937)	-0.332 (1.388)	0.937	0.032
Encouragement of Cognitive Problem-Solving Strategies	0.211 (2.211)	0.486 (1.326)	0.660	-0.180
Use of Pedagogically Based Learning Strategies	-0.397 (2.084)	-0.172 (0.890)	0.639	-0.192

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from independent-samples *t* tests; they are provided for parallelism with the outcome analysis results.

Baseline (In)equivalence on Teacher and AP CSP Student Characteristics

Initially we had hoped to be able to include some teacher and classroom characteristics as matching variables, but initial testing with the full eligible pool from the 2020 data (11 Lone Star teachers and 87 potential comparison teachers) indicated that we would not be able to achieve good matches on the background characteristics if we were to achieve good matches on the instructional practice measures (and vice versa). As shown in Appendix Exhibits C-2 though C-5 for the Year 1 sample, Appendix Exhibits C-6 though C-9 for the Year 2 sample, and Appendix Exhibits C-10 through C-13 for the Year 3 sample, there were some notable baseline differences between the Lone Star teachers and the comparison teachers, as well as between the AP CSP students taught by each group of teachers.

In particular, the Lone Star teachers tended to be less experienced and younger than the comparison teachers, and they were less likely to be female; in the Year 1 sample, they were also more likely to report holding a master's degree in a STEM field. In terms of the student composition of the 2019–20 (baseline) AP CSP classes taught by the teachers, in all three years' samples, the Lone Star teachers reported much higher percentages of Hispanic (or Latinx) students and much lower percentages of White students in their AP CSP classes than did the comparison teachers. Similarly, they reported higher percentages of English language learner students. In addition, especially in the Year 1 and Year 3 samples, the Lone Star teachers reported higher proportions of ninth-grade students and lower proportions of 12th-grade students.

Despite these differences, we elected to include only the three instructional practice measures as matching variables, given the greater importance of baseline equivalence on the outcome measures than on background characteristics, and the inability to achieve baseline equivalence

on both types of variables given the pools of teachers on whom we had data. The extent to which our impact analysis strategy accounted for differences in background characteristics is described toward the end of the following section.

Statistical Estimation Strategy

To estimate the effect of the Lone Star program on teachers' instructional practices, we ran a separate ordinary least squares (OLS) regression for each of the three outcomes in each of the three years, accounting for teachers' baseline values on that outcome. The unit of analysis was the teacher, and given that all (or nearly all) teachers were in different schools (and thus the teacher level and the school level were conflated), there was no need for a multilevel structure. The analytic model had the following general specification:

$$y_s = \beta_0 + \beta_{Trt} Lonestar_s + \beta_{Pre} PreOutcome_s + \boldsymbol{x}_s' \boldsymbol{\beta}_{\boldsymbol{Cov}} + \varepsilon_s$$

where y_s represents the outcome of interest of teacher *s*, $Lonestar_s$ is a binary variable indicating whether teacher s is a Lone Star teacher, and $PreOutcome_s$ represents the pre-intervention (i.e., baseline) measure of the outcome. The vector $\boldsymbol{x}_s'$ represents teacher-level characteristics (see below), and ε_s is the error term. The coefficient β_{Trt} represents the estimated intervention effect.

After conducting this regression analysis, we calculated the effect size of the estimated intervention effect for each of the three measures. Again we used the formula for Hedges' *g* as described earlier, except that instead of the numerator being the *un*adjusted mean difference between the Lone Star and comparison groups at baseline, it was the *adjusted* mean difference in the *outcome* year, namely the β_{Trt} coefficient from the regression. (The standard deviations in the denominator were also from the outcome year.)

Covariate Selection

To select the teacher characteristics to be included as covariates in the $\boldsymbol{x}_s'$ term, we modeled the 2020 (baseline) scores for each of the three outcomes on several different characteristics for which we had data, all of which were self-reported on the teacher survey. We employed this modeling process separately for each of the three analysis samples previously mentioned. Because the sample sizes were quite small, we tested each characteristic individually, with no other variables in the model, i.e.:

$$Y_{\text{PreOutcome}} = \beta_0 + \beta_1 \text{Characteristic}_s + \varepsilon_s$$

The characteristics we tested were:

- Total years of teaching experience (any subject, any grade level)[6]
- Years of experience teaching computer science[7]
- Whether the teacher had a master's degree in a STEM field
- Whether the teacher was certified to teach computer science
- Whether the teacher was certified to teach mathematics
- Whether the teacher had ever held a full-time job in computer science that was not related to teaching or education
- The number of years the teacher had held a full-time job in computer science not related to teaching or education (coded as 0 for teachers who had never held such a job)
- Whether the teacher was female
- The teacher's age

Our pre-established decision rule was to retain for the impact analysis any characteristic that had a *p* value of less than .10—a rather low threshold, because of the need for parsimony given the small sample sizes—in this exploratory baseline model for any of the three instructional practice measures. We also decided that any characteristic meeting this criterion for at least one of the three practice measures in a given year would be included in the impact analysis model for all three measures for that year, so as to have a common set of covariates. However, given that the teacher samples for the Year 1, Year 2, and Year 3 analyses differed, we did not require inclusion of the same set of covariates across the three years.

Teacher Characteristics Selected for Inclusion as Covariates

In the sample of teachers for the Year 1 analysis, the covariate selection process yielded three teacher characteristics to be included in the impact model:

- Whether the teacher had a master's degree in a STEM field, which negatively predicted the baseline value of the first measure, Engaging Students in Practices of Computer Science ($p < .10$)
- The number of years the teacher had held a full-time job in computer science, which negatively predicted the baseline values of all three measures ($p < .05$ for Engaging

[6] Total years of teaching experience was a continuous variable. Very few teachers in either year's sample had fewer than four years of experience, so there was no need for a "novice vs. veteran" dichotomous specification.

[7] We tested three different specifications of years of computer science teaching experience. One was as a continuous variable. The other two were dichotomous: one for whether the teacher had more than two years of computer science teaching experience as of the 2019–20 school year, and the other for whether the teacher had more than three years of computer science teaching experience as of the 2019–20 school year.

Students in Practices of Computer Science and Encouragement of Cognitive Problem-Solving Strategies; $p < .10$ for Use of Pedagogically Based Learning Strategies)

- Whether the teacher was female, which positively predicted the baseline value of the third measure, Use of Pedagogically Based Learning Strategies ($p < .10$)

In the Year 2 analysis sample, the covariate selection process yielded just a single teacher characteristic to be included in the impact model: the number of years the teacher had held a full-time job in computer science. This negatively predicted the baseline values of each of the first two instructional practice measures, Engaging Students in Practices of Computer Science and Encouragement of Cognitive Problem-Solving Strategies ($p < .10$ for each).

In the Year 3 analysis sample, the process resulted in the selection of two covariates for the impact analysis: the number of years the teacher had taught computer science and, as in the previous two years' samples, the number of years the teacher had held a full-time job in computer science. Each of these predicted only the first instructional practice measure, Engaging Students in Practices of Computer Science—the former positively and the latter negatively ($p < .05$ for each). None of the teacher characteristics we tested predicted either the second or third practice measures.

As already noted, teacher's baseline (2020) score on each practice measure was also included as a covariate in the impact model for that measure. None of the covariates included in any of the three years had any missing data.

A Note About Classroom (and Student) Characteristics

We did not include any classroom or student characteristics in our analysis models for several reasons. The first has to do with the fact that these characteristics change from year to year. It would not make much sense to model instructional outcomes for a given school year (2020–21, 2021–22, or 2022–23) adjusting for classroom composition from an earlier year (2019–20). On the other hand, although it might seem intuitive to adjust for classroom characteristics in the *same* year as the instructional outcomes, this is methodologically problematic given the possibility that the Lone Star intervention could itself have an effect on classroom composition. (Indeed, the program aims to increase access to AP CSP, particularly for underrepresented minority students.) Controlling for variables that might be affected by the intervention being studied, sometimes called *overadjustment*, can cause bias in the estimation of the effect of the intervention and thus is generally discouraged (e.g., Groenwald et al., 2019; VanderWeele, 2019; Williams et al., 2018); typically, one should only include covariates measured prior to the start of the intervention (i.e., "pretreatment").

Second, some teachers did not answer the questions on the student composition of their AP CSP classes, creating a problem of *missing data*. A standard method of handling missing data in a covariate is to include in the analysis model an additional variable indicating missingness of data on that covariate (called "dummy variable imputation"); however, given our very small sample sizes, inclusion of additional variables increased the possibility that our models would be overspecified and also potentially decreased their statistical power. Indeed, there was a need to be as parsimonious as possible in inclusion of covariates, which is another reason why we were loath to include classroom characteristics.

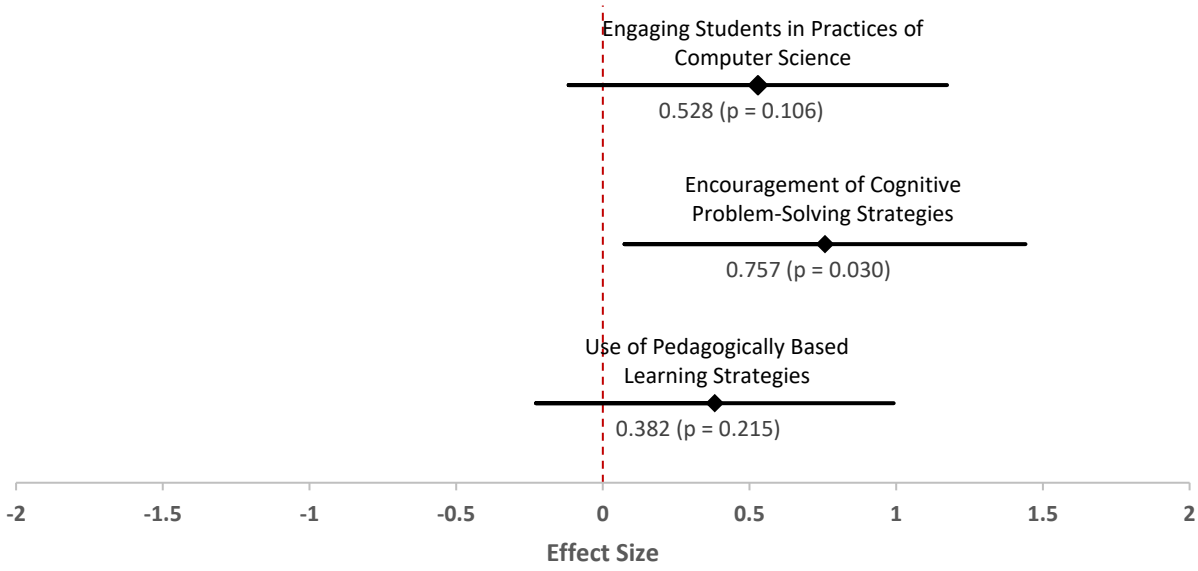
Finally, in analyses using data only for the baseline (2019–20) year, per the analyses described just above regarding selection of teacher characteristics to use as covariates, most of the classroom composition variables on which we had data (see Appendix Exhibits C-4, C-5, C-8, C-9, C-12, and C-13) did not significantly predict any of the three instructional practice outcomes. This provided reassurance that classroom composition variables were not needed as controls in our impact analyses.

Results

Exhibits 4.7, 4.8, and 4.9 show the estimated effect sizes for the effects of the Lone Star AP CSP program on the three instructional practice measures after one, two, and three years of implementation, respectively. Exhibit 4.10 provides additional details, including the unadjusted (i.e., descriptive) means and standard deviations of the three measures in each outcome year. The full regression results are provided in the Appendix (Exhibits C-14, C-15, and C-16 for each year, respectively).

As shown in Exhibit 4.7, after being implemented for one year, the Lone Star program had a significant positive effect on one of the three instructional practice measures: Encouragement of Cognitive Problem-Solving Strategies. This effect was about three quarters of a standard deviation, a sizeable effect by most standards. The one-year effects of the program on the other two measures—Engaging Students in Practices of Computer Science and Use of Pedagogically Based Learning Strategies—also were positive, but were not statistically significant, which means that we cannot, with confidence, rule out that the true effect (as opposed to the estimated effect) was actually null, or zero.

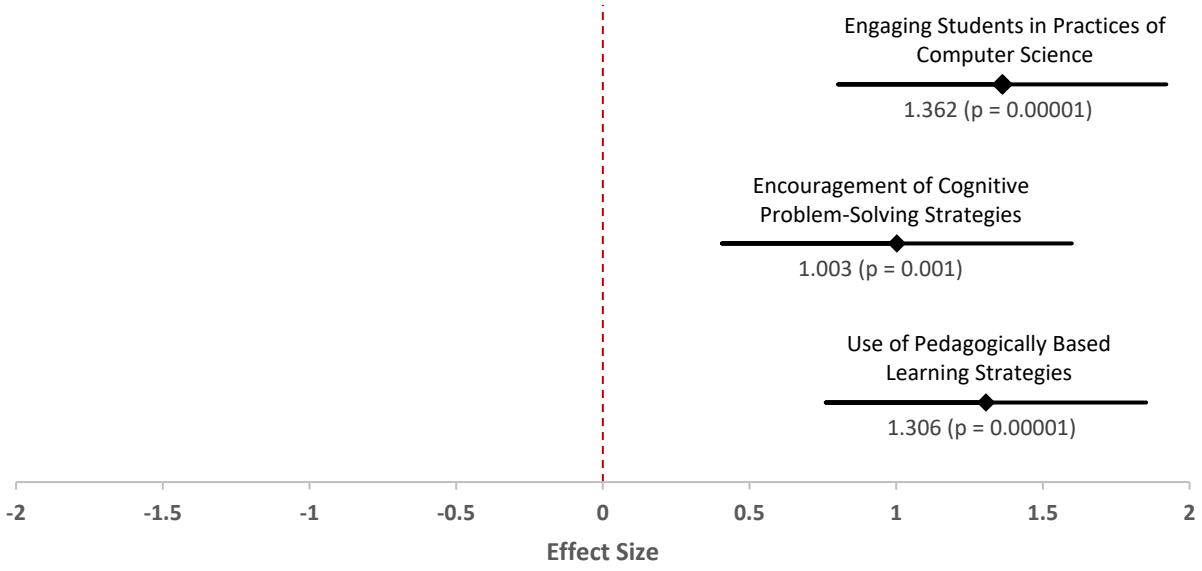
Exhibit 4.7. Estimated Effects of Lone Star AP CSP on Teachers' Instructional Practices After One Year of Implementation



Note. *N* = 63 (9 Lone Star teachers and 54 comparison teachers). Effect size point estimates are represented by the diamonds; the horizontal lines show 95% confidence intervals. Confidence intervals that do not intersect with an effect size of 0 (represented by the dotted red line) indicate a statistically significant ($p < .05$) effect.

After two years of Lone Star implementation, the program had a significant positive effect on all three of the instructional practice measures (Exhibit 4.8). Moreover, all three estimated effects were very large by nearly any standard, at more than a full standard deviation—in fact, 1⅓ standard deviations for two of the three measures.

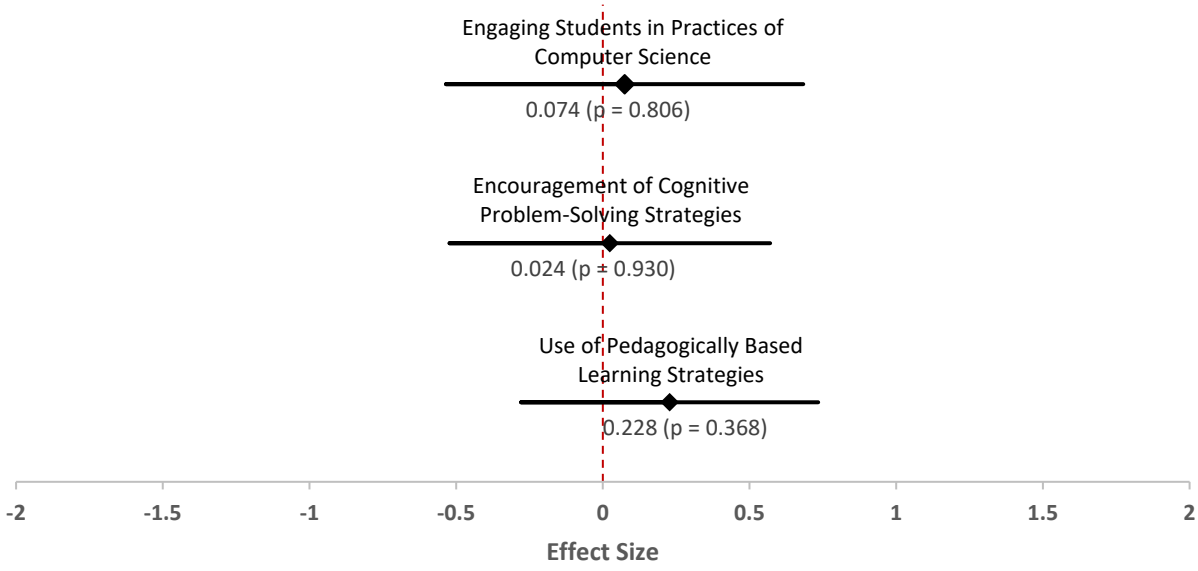
Exhibit 4.8. Estimated Effects of Lone Star AP CSP on Teachers' Instructional Practices After Two Years of Implementation



Note. *N* = 56 (7 Lone Star teachers and 49 comparison teachers). Effect size point estimates are represented by the diamonds; the horizontal lines show 95% confidence intervals. Confidence intervals that do not intersect with an effect size of 0 (represented by the dotted red line) indicate a statistically significant ($p < .05$) effect.

After three years of Lone Star implementation, the program did not have a significant positive effect on any of the three instructional practice measures (Exhibit 4.9). Not only were the estimated effect sizes statistically insignificant, they were also very small. The first two measures, in particular, had an effect size that was essentially 0, with nearly as much of the confidence interval being negative as positive. Therefore, the lack of significance in the three-year analysis cannot be attributed simply to the fact that the sample was smaller in this year than in the previous two years.

Exhibit 4.9. Estimated Effects of Lone Star AP CSP on Teachers' Instructional Practices After Three Years of Implementation



Note. *N* = 42 (7 Lone Star teachers and 35 comparison teachers). Effect size point estimates are represented by the diamonds; the horizontal lines show 95% confidence intervals. Confidence intervals that do not intersect with an effect size of 0 (represented by the dotted red line) indicate a statistically significant ($p < .05$) effect.

Exhibit 4.10. Outcome Statistics for All Three Years' Samples

Measure	Lone Star teachers (unadj.)	Comparison teachers (unadj.)	Adj. mean difference (β_{Trt})	*p* value	Hedges' *g*
After one year (9 Lone Star teachers and 54 comparison teachers)					
Engaging Students in Practices of Computer Science	0.227 (1.090)	-0.556 (1.265)	0.6650 (0.4053)	0.106	0.528
Encouragement of Cognitive Problem-Solving Strategies	1.601 (1.903)	0.552 (1.649)	1.2913 (0.5819)	0.030	0.757
Use of Pedagogically Based Learning Strategies	-0.368 (1.310)	-0.791 (1.260)	0.4897 (0.3904)	0.215	0.382
After two years (7 Lone Star teachers and 49 comparison teachers)					
Engaging Students in Practices of Computer Science	2.289 (3.222)	0.076 (1.149)	2.1064 (0.4310)	0.000	1.362
Encouragement of Cognitive Problem-Solving Strategies	2.791 (3.310)	1.157 (1.261)	1.6493 (0.4888)	0.001	1.003
Use of Pedagogically Based Learning Strategies	1.439 (2.423)	-0.048 (0.970)	1.6162 (0.3368)	0.000	1.306
After three years (7 Lone Star teachers and 35 comparison teachers)					
Engaging Students in Practices of Computer Science	0.019 (1.749)	-0.211 (1.124)	0.0939 (0.3792)	0.806	0.074
Encouragement of Cognitive Problem-Solving Strategies	0.736 (1.702)	0.581 (1.354)	0.0344 (0.3878)	0.930	0.024
Use of Pedagogically Based Learning Strategies	-0.209 (1.487)	-0.425 (0.959)	0.2449 (0.2686)	0.368	0.228

Note. In the first two columns, each cell contains the unadjusted mean and (in parentheses) standard deviation of the outcome measure. The third column contains the regression coefficient on the Lone Star indicator from the impact analysis model (i.e., the adjusted mean difference) and (in parentheses) the standard error of the coefficient; the *p* value associated with the coefficient is in the fourth column. The Hedges' *g* value in the final column results from the inclusion of the regression coefficient as the numerator (i.e., the mean difference) in the Hedges' *g* formula.

As shown in Appendix Exhibits C-14 through C-16, the baseline measure of the outcome was a highly significant predictor of all three instructional practice measures in all three years. On the other hand, the teacher covariates we included were generally not significant.

Exploration of Differences Between the Two- and Three-Year Effects

That the very large (and statistically significant) positive effects after two years of implementation were followed by essentially null effects after three years of implementation is head-scratching, to say the least. We explored a few possible reasons for the discrepancy.

We began by checking into and ruling out explanations related to survey administration. To our knowledge, there were no major differences between the 2022 and 2023 survey administrations in the communications to the teachers about the survey or its purposes, although as previously noted, NMSI handled direct contact with the Lone Star teachers. There also were no changes to the relevant survey questions, the survey platform, the user interface, or the procedures used to clean and process the survey data. (And even if there had been, the changes would have affected Lone Star teachers and comparison teachers alike, so could not have explained the impact difference.)

Sample differences represented another possible reason for the differences in the two years' results. However, given that the two Lone Star samples overlapped by six of seven teachers (see Exhibit 4.3 above), sample differences did not seem likely to explain the pattern. Nor did outliers in one or both years. In fact, of the six Lone Star teachers who were in both years' analyses:

- Every one of the six had a decline in their score on Engaging Students in Practices of Computer Science (i.e., indicating less frequent use of the practices in 2022–23 than in 2021–22).
- Five of the six had a decline in their score on Encouragement of Cognitive Problem-Solving Strategies. (The sixth teacher had an increase.)
- Five of the six had a decline in their score on Use of Pedagogically Based Learning Strategies. (The sixth teacher had no change.)

Relatedly, the difference between the two years' effects is not attributable to large increases among the comparison teachers (relative to the changes among the Lone Star teachers). In fact, as shown in Exhibit 4.10 above, the comparison teachers' scores on all three measures also dropped from the second outcome year to the third outcome year—just not by as much as the Lone Star teachers' scores did.

Finally, we examined program usage patterns among the six Lone Star teachers who were in both the second- and third-year analyses. Of the Lone Star program's six main activities for teachers in Years 2 and 3 (see Exhibit 3.2 in Chapter 3 on program usage), usage dropped slightly among the six teachers, from an average of 4.5 activities in Year 2 down to an average of 4.0 in Year 3. Three of the teachers participated in the same number of activities in the two

years; the other three all participated in one fewer activity in Year 3 than in Year 2. For the program activities that the teachers *did* participate in in either year, average usefulness ratings among the six teachers also dropped a bit, from 2.9 (just below *moderately useful*) in Year 2 to 2.5 (halfway between *slightly useful* and *moderately useful*) in Year 3. Therefore, it is possible that lower participation levels in the program and lower perceptions of program activities' usefulness contributed to the declines observed in the instructional practice measures. However, the participation and perception declines do not seem large enough to fully account for the large drops in the instructional practices.

Therefore, unfortunately, we are unable to fully explain why the program effects on instructional practices declined so dramatically from the second year (the 2021–22 school year) to the third year (the 2022–23 school year). NMSI did not inform us of any major changes to the program from the second to third year that might have caused the decline. We also are not aware of any changes to the national or Texas state context for teaching AP CSP, or for teaching high school more generally, that could have interacted with the program to cause a larger decline for Lone Star teachers than for the matched comparison teachers. Perhaps in 2021–22, when schools returned to being fully open for in-person instruction after the largely virtual 2020–21 year, teachers experienced a renewed sense of energy and enthusiasm—which could have translated into willingness to try new and different instructional practices—that then wore off in the 2022–23 year as it began to seem that pandemic effects on students were lingering. But again, it is not clear why Lone Star AP CSP teachers would have experienced greater year-to-year swings in their use of particular instructional practices than the comparison teachers.

Summary and Discussion

In summary, we found that the Lone Star AP CSP program had positive effects on participating teachers' instructional practices after the first and second years of implementation but not after the third. Specifically, we observed:

- a significant positive effect after one year of implementation on one of the three instructional practice measures we examined: Encouragement of Cognitive Problem-Solving Strategies
- a large and significant positive effect after two years of implementation on all three of the instructional practice measures:
  - Engaging Students in Practices of Computer Science
  - Encouragement of Cognitive Problem-Solving Strategies
  - Use of Pedagogically Based Learning Strategies
- no effect after three years of implementation on any of the three measures

We are not able to fully explain the pattern of results across the three years—suggestive findings in the first year, large positive effects in the second year, and no effects in the third year.

That said, it may be useful to consider the instructional practice findings in the context of the implementation, uptake, and participant engagement findings discussed in previous chapters of this report. As noted in Chapters 2 and 3, the first year of the Lone Star program was the first full year of the COVID-19 pandemic, which could have negatively affected teacher engagement in and uptake of the program that year, especially given that some professional development offerings that had originally been envisioned as in-person opportunities became virtual. By some metrics, teacher uptake did increase in Year 2, but whether the somewhat higher level of participation in Year 2 was sufficient to account for the very large effects on instructional practice observed that year is debatable. Even more debatable is whether changes to participation and engagement levels from Year 2 to Year 3 can explain the complete loss of instructional practice effects in Year 3. The data presented in the previous chapters indicate relatively little change from Year 2 to Year 3 in implementation of the teacher component and in teacher participation levels. When we look solely at the six Lone Star teachers common to the Years 2 and 3 instructional practice impact analyses, we find that their participation and engagement were lower in Year 3 than in Year 2, but not dramatically lower.

A number of caveats should be kept in mind in interpretation of the results. Perhaps most importantly, the measures of teachers' instructional practice are based on the teachers' own self-report on retrospective year-end surveys; we did not ask teachers to report on or log their practices over the course of the year, nor did we independently observe teachers' instruction. It is possible that teachers may have misreported or misrepresented their instructional practices (whether intentionally or not) as they reflected, while completing a one-shot survey, at or even after the end of the school year.

The small size of the sample—particularly the very small number of Lone Star teachers included—is another major limitation of the study. Statistical estimates vary more with small samples; such variation also may help explain the changing pattern of results across the three years. In addition, the study did not randomly assign teachers (or schools) to the program. It is a strength of our design that the comparison teachers were *not* in schools that were offered the program but turned it down, as that would automatically make them different from the Lone Star teachers in ways that could clearly influence the outcomes. However, we do not know whether the comparison teachers (or their schools) would have accepted the program *if* they had been offered it—most likely, some would have and some would not have. (We do know that all of the comparison teachers were willing to participate in the *evaluation*, but completing surveys is less burdensome than participating in professional development.)

Other limitations are as follows:

- We cannot assume that the effects (or lack thereof) observed in our samples would generalize to a larger or different set of teachers. We already know that the Lone Star teachers included in the analyses are not representative of all the Lone Star teachers, in that they, unlike most of the excluded teachers, taught AP CSP in the 2019–20 school year.
- The small sample size limited the number of covariates we could include in the analysis. In addition, our analysis approach and estimation strategy did not allow for the inclusion of time-varying classroom characteristics as controls, given that they could be affected by the intervention. Although preliminary analyses indicated that teacher and classroom characteristics were generally not significantly associated with the instructional practices we examined, we cannot entirely rule out a relationship. And on some of the characteristics, there were clear differences between the Lone Star teachers and the comparison teachers. Taken together, these background characteristic differences between the groups could help explain estimated differences in use of instructional practices.
- The instructional practice measures are all measures of the *frequency* with which teachers used particular instructional practices, not the *quality*; just because a practice is used often does not mean it is used well.

The study limitations apply equally to all three years' analyses. They are enough to warrant caution in interpreting both the large positive effects found after the second year and the null effects found after the third year. To the extent that there is a link between teachers' instructional practices and student outcomes on the AP CSP exam, the following chapters—about the effects of the Lone Star program on AP CSP exam performance and participation—may help tell more of the story. In particular, if the pattern of findings of null or small positive effects in Year 1 followed by large positive effects in Year 2 followed by null effects in Year 3 is mirrored in Lone Star schools' AP CSP exam performance over the three years, we can have greater confidence in the instructional practice findings.

Chapter 5. Impact on Student AP CSP Performance and Participation After One Year

As detailed in the final column of the program logic model (Exhibit 2.1 in Chapter 2), the ultimate aims of the Lone Star AP CSP program were to increase the likelihood that (a) AP CSP course takers succeed in the course, and (b) high school students take the AP CSP course and exam. This chapter examines whether the program achieved these aims during its first year of operation, the 2020–21 school year. The research question addressed is as follows:

4. What was the impact of one year of participation in the Lone Star AP CSP program, compared to business-as-usual AP CSP participation, on high schools' rates of (a) AP CSP exam takers receiving a qualifying score on the AP CSP exam, and (b) students taking the AP CSP exam?

The following chapter will look at the impact after two years of implementation, and the chapter after that will look at the impact after three years of implementation.

As in the analysis of instructional practice, the student impact study used a quasi-experimental matched comparison group design and an analysis-of-covariance (regression) estimation model to examine the effect of Lone Star AP CSP on the performance and participation outcomes. We conducted separate regressions for each of two samples of schools (those that had AP CSP exam takers in the baseline year and those that did not) and then conducted meta-analysis to combine the estimates.

KEY FINDINGS

- After one year of implementation, the Lone Star AP CSP program did not have a significant effect on schools' percentages of AP CSP exam takers that received a qualifying score on the exam.
- The program also did not have a significant effect after one year on schools' percentages of students in grades 9–12 that received a qualifying score—or a nonqualifying score—on the AP CSP exam.
- There was also no significant effect after one year of implementation on schools' percentages of students in grades 9–12 that took the AP CSP exam.

The remainder of this chapter is organized as follows. First, we describe and discuss the measures of AP CSP achievement and participation we examined as outcomes, along with our sources of data on these measures. We then discuss the sample of Lone Star schools included in the analysis and explain how we identified comparison schools that were similar at baseline to the Lone Star schools.

We then detail our statistical estimation strategy, followed by presentation of the results. The chapter concludes with a brief summary and discussion.

Outcome Measures

The outcome measures for the student impact study fall into two categories: performance and participation. Performance pertains to how well students did in AP CSP, which was theorized to be affected by Lone Star AP CSP mainly through the supports for teachers and the supports for students, and participation pertains to how many students took AP CSP, which was theorized to be affected mainly by the supports for schools. We discuss each category in turn.

Performance in AP CSP

As indicated by the phrasing of the research question, AIR's evaluation operationalized success in the AP CSP course as earning a qualifying score—that is, a 3, 4, or 5—on the AP CSP exam.[1] Scores of 3 and higher are called "qualifying" scores because at many colleges and universities, receiving such a score qualifies a student for college credit and/or placement out of an introductory course (College Board, n.d.-b.). AP exam score is therefore a meaningful measure for students and has clear policy relevance. In addition, past research has found that scores of 3 or higher are predictive of college success (Dougherty, Mellor, & Jian, 2006; Morgan & Klaric, 2007).

Student-level data were not available to the research team given the quasi-experimental design of the study (involving Lone Star AP CSP schools in multiple school districts or charter management organizations and matched comparison schools elsewhere with which the evaluation did not expect to have any contact).[2] Therefore, we designated our unit of analysis as the school, making the student impact study a cluster QED, with the outcomes being percentages of students rather than individual student scores. Our key measure of the performance outcome, per the program logic model, was the *percentage of AP CSP exam takers earning a 3, 4, or 5 on the exam.* That is, of the students at a school *who took the AP CSP exam,* what percentage of them scored 3 or above?

[1] Administered by the College Board each spring, AP exams are scored on an ordinal basis from 1 (lowest score) to 5 (highest score).

[2] The College Board's policies for sharing data with researchers are quite strict (College Board, 2018). This is true even for institution-level data, and even more so for individual-level data, even if deidentified. To obtain student-level data, AIR likely would have needed explicit agreements with all of the local education agencies (LEAs) that had schools selected for inclusion in the analyses, especially if we wanted to affix other student-level data to the AP exam data, such as student background characteristics or achievement on other measures. This would have been challenging even just with the 15 LEAs in which the Lone Star AP CSP schools were located. For comparison schools, even if we had still *matched* at the school level using cohort-year institution-level AP data obtained from the College Board, as we actually did, subsequently obtaining student-level data through agreements with all of the LEAs that had schools selected as matches—numbering well over 100 LEAs for each year's analysis, and with differing schools and LEAs across the three years given that not all schools had AP CSP exam takers every year—would have been both time- and cost-prohibitive. It was always the intent of the evaluation, even at the proposal stage, to use the school as the unit of analysis.

To improve the likelihood that the student impact study meets WWC criteria for cluster QEDs, we supplemented this outcome measure with two additional measures: the percentage of students *in grades 9–12*[3] earning a 3, 4, or 5 on the AP CSP exam and the percentage of students *in grades 9–12 earning a 1 or 2* on the AP CSP exam. Only the first of these two additional measures would be needed to meet WWC criteria, but the second is an important complement to the first, because if the denominator for the percentage calculations is the entire enrollment of grades 9–12 (rather than just the students who take the AP CSP exam), a school could, in theory, have increases in both the percentage of students scoring a 3 or above *and* the percentage of students scoring 2 or below (see the box for an example). The former increase is desirable; the latter increase is not. That is, the goal would be to increase the numbers and percentages of students succeeding in the course *without* simultaneously increasing the numbers and percentages of students doing poorly in the course. (This is, presumably, why Lone Star AP CSP focused on professional development for AP CSP teachers and supports for AP CSP students, as well as on helping schools identify more students with the potential to succeed in the course and then encouraging those students—not just anyone—to take the course.) Thus, we consider it essential to include the second supplementary outcome if including the first supplementary outcome.

ILLUSTRATION OF WHY THE SECOND SUPPLEMENTARY MEASURE IS NEEDED IF THE FIRST SUPPLEMENTARY MEASURE IS INCLUDED

Imagine a school with a constant enrollment of 100 students in grades 9–12. If, in the baseline year, 10 students took the AP CSP course and exam, and seven of them received a qualifying score and the other three did not, the percentage of exam takers earning a qualifying score is 70%, the percentage of grades 9–12 students earning a qualifying score is 7%, and the percentage of grades 9–12 students earning a nonqualifying score is 3%. If, in an implementation year, 20 students took the AP CSP course and exam, and 11 of them received a qualifying score and the other nine did not, the percentage of exam takers earning a qualifying score is 55% (a decrease from the baseline 70%), the percentage of grades 9–12 students earning a qualifying score is 11% (an increase from the baseline 7%), and the percentage of grades 9–12 students earning a nonqualifying score is 9% (also an increase from the baseline 3%). Thus, the school has increased the percentage of all grades 9–12 students earning a qualifying score on the AP CSP exam, but that increase is largely a function of having increased the percentage of students *taking* the exam, such that the school also increased the percentage of students receiving a nonqualifying score. Thus, it would be an incomplete picture to look only at the percentage of grades 9–12 students earning a qualifying score.

[3] We included only grades 9 through 12 rather than whole-school enrollment because the AP program targets the high school grades. Most schools with AP courses are high schools that only have grades 9–12, but some do have other grade levels.

The program's theory would posit a positive effect on the first and a negative effect on the second.

Participation in AP CSP

As with student-level data, data on course enrollments were not available for the evaluation. Because of this, along with the linkage between participation and performance just discussed, the evaluation operationalized participation in the course as taking the AP CSP exam. Although there might be some course takers who do not take the exam,[4] it could be argued that a substantial portion of the value of an AP course lies in the AP exam, in terms of college placement or credit, and thus that taking the exam is a key element of course participation. In fact, research suggests positive relationships between AP *exam taking* and college outcomes (Mattern, Shaw, & Xiong, 2009; Mattern, Shaw, & Ewing, 2011; Mattern, Marini, & Shaw, 2013; Murphy & Dodd, 2009). Accordingly, the key outcome for the second part of the research question, about participation, was the percentage of students in grades 9–12 taking the AP CSP exam.

Summary of Outcome Measures

In summary, we examine four outcomes in the analysis of the impact of the Lone Star AP CSP program on student AP CSP performance and participation:

- the percentage of AP CSP exam takers earning a 3, 4, or 5 on the exam *(key measure of student performance)*;
- the percentage of students in grades 9–12 earning a 3, 4, or 5 on the AP CSP exam *(supplementary measure #1 of student performance)*;
- the percentage of students in grades 9–12 earning a 1 or 2 on the AP CSP exam *(supplementary measure #2 of student performance)*; and
- the percentage of students in grades 9–12 taking the AP CSP exam *(measure of student participation)*.

[4] The AP CSP teacher survey we administered annually from 2020 through 2023 (see Chapters 3 and 4) asked the surveyed teachers to provide the approximate percentage of students in their AP CSP course who had taken that year's AP CSP exam. Results are presented in Appendix Exhibit D-9. The exhibit shows that, according to the respondents' self-report, exam-taking was lowest in spring 2020; only 18% of respondents reported that nearly all (i.e., 95 to 100%) of their AP CSP students submitted the digital portfolio that constituted that year's AP CSP exam. (This is not surprising given that spring 2020 was when the COVID-19 pandemic began; unlike many other standardized tests, AP exams were not canceled entirely, but obviously the disruptions were severe.) Exam-taking rates were still low in 2021, but by 2022 and 2023, the rates had risen considerably, exceeding what they had been in pre-pandemic 2019. Notably, in both 2022 and 2023, about half of the responding teachers reported that nearly all of their AP CSP students took the AP CSP End-of-Course exam. However, we do not know how representative (or even accurate) these results are, given that they are based on teacher self-report from a nonrepresentative sample of teachers, most of whom did *not* teach at schools included in the AP impact analyses presented in this chapter.

MATHEMATICAL RELATIONSHIPS AMONG THE FOUR OUTCOME MEASURES

At the level of the individual school, the four outcomes are mathematically related to one another. For instance, if you know the middle two for a given school, you can calculate the first and fourth for that school. However, when averaging over multiple schools, the mathematical relationships are much less straightforward, particularly for the first outcome, and they become even more complex when taking into account variability around the means, other factors that might affect the outcomes, and so forth, as we do in statistical analysis. The first outcome—percentage of passers among takers—and the fourth outcome—percentage of students in the school taking the exam—are numerically independent of one another, and they thus could be analyzed as the only two outcomes, with the latter as the measure of performance and the former as the measure of participation. However, as noted, we added the other two based on What Works Clearinghouse standards for cluster QEDs.

For the analysis of the impacts of the first year of Lone Star AP CSP—the 2020–21 school year—all of these outcome percentages were based on the AP CSP exam administered in spring 2021. For the three measures using grades 9–12 enrollment as the denominator for the percentage calculations, school enrollments were from the same school year (2020–21).

Data Sources for the Outcome and Baseline Measures

Data sources differed depending on the state in which schools were located. (As shall be discussed below, although all of the Lone Star AP CSP schools were in Texas, we allowed comparison schools to be in other states as well as in Texas.)

Texas schools. AP data for Texas schools, including the Lone Star schools, came from datasets prepared by the College Board for the AIR evaluation team, with the approval of the Texas Education Agency, for each of exam years 2020 (the pre-program baseline year), 2021 (the outcome year studied in this chapter), 2022 (the outcome year studied in Chapter 6), and 2023 (the outcome year studied in Chapter 7). Each year's file contained AP CSP exam data for every Texas public school that had any AP CSP exam takers that year, including the number of students who took the AP CSP exam, and, if there were at least five AP CSP exam takers, the number of students receiving each of the five possible scores (1 to 5) on the AP CSP exam. The data for 2020 also included every Texas public school with AP exam takers in any AP subject area,[5] including the number of exams taken across all subject areas and, if there were at least five exams total, the number of exams (again, across all subject areas) receiving each score.

[5] Throughout the study period (from the 2019–20 school year through the 2022–23 school year, there were 38 AP subject areas, each with its own AP exam.

Other states. To provide a larger pool of comparison schools within which to match Lone Star schools for the analysis, the evaluation team supplemented the College Board Texas-only files with data from other states.[6] Three states had publicly available school-level, subject-specific AP exam data that could be used to supplement the Texas data: Georgia, Massachusetts, and Wisconsin.[7] In addition, New York had publicly available district-level AP exam data,[8] which we were able to "convert" to school-level data for districts having only a single high school, given that the AP program is typically found only at the high school level. This resulted in AP data for more than 400 New York schools[9]—about 86% of the districts with AP exam takers in the state.

Like the Texas data, the AP exam data for all four of these other states did have some data suppression based on numbers of exam takers, but suitably high numbers of schools in these states had usable data.

Enrollment data. Because three of the evaluation's four outcome measures are based on percentages of students in grades 9–12, we also needed grade-level enrollment data for each school. We obtained these enrollment data from the websites of the five states[10] and merged them into the AP data so that we could calculate the percentages. This merging was relatively straightforward for the four non-Texas states where both the AP data and the enrollment data came from state websites and schools had a common identifier (state school ID) across files. It was more difficult for Texas, where the school ID numbers in the AP data were College Board IDs. Between publicly available Texas files for some school-level AP data (not subject-specific and only for students in grades 11 and 12) that had both College Board IDs and state school IDs for many (though not all) of the schools in the College Board data,[11] a crosswalk between College Board IDs and federal IDs provided by the College Board, and manual lookups in federal,

[6] Because the AP program is national and not strongly affected by state and local policy or conditions, we considered it legitimate for *some* (but not all) comparison schools to be located outside of Texas. Although one might think that Texas, as one of the nation's largest states, would offer a sufficiently large pool of comparison schools from which to select good matches for the Lone Star AP CSP schools, we were concerned that it might not. High schools are not as numerous as lower-level schools; furthermore, relatively few high schools offer AP CSP—a course that has formally existed only since the 2016–17 school year—and schools that offer the course in one year might not offer it in other years. In the 2020 (baseline year) data we received from the College Board, a total of 325 Texas high schools had AP CSP exam takers, but only 252 of them had unsuppressed AP CSP exam data. Moreover, in a Texas data file that NMSI had access to early on in the evaluation, which informed some of our decisions, the suppression rules were much more severe—a school needed to have at least 10 male exam takers and 10 female exam takers for its score data to be unsuppressed—such that only 64 Texas schools had unsuppressed AP CSP exam data in 2020, and we knew that not all of these 64 schools would have unsuppressed AP CSP exam data in any given outcome year. On this basis (before we knew that the Texas AP data we would obtain would have less suppression) we decided to allow comparison schools in other states.

[7] https://gosa.georgia.gov/dashboards-data-report-card/downloadable-data, https://profiles.doe.mass.edu/statereport/ap.aspx , and https://dpi.wi.gov/wisedash/download-files, respectively

[8] https://data.nysed.gov/downloads.php

[9] 455 in 2020, 420 in 2021, 444 in 2022, and 439 in 2023.

[10] Same websites as for the AP data for Georgia, New York, and Wisconsin. For Texas, https://rptsvr1.tea.texas.gov/adhocrpt/adste.html and for Massachusetts, https://www.doe.mass.edu/infoservices/reports/enroll/default.html.

[11] https://tea.texas.gov/reports-and-data/school-performance/accountability-research/advanced-placement-and-international-baccalaureate

Texas, and College Board school directories,[12] we were able to merge the Texas AP data with the Texas school enrollment data and calculate percentages for nearly all of the schools.

Lone Star School Sample and Selection of Comparison Schools

As noted in earlier chapters, 37 Texas schools were recruited by NMSI to participate in the Lone Star AP CSP intervention. One of the 37 schools was excluded from all AP exam impact analyses (in all outcome years) because it did not have any AP exam takers (of any subject) in 2020 and our identification of comparison schools—as well as our determination of baseline equivalence—was based on baseline (2020) AP exam taking and scores.

Of the remaining 36 Lone Star schools, 28 were included in the analyses of the impact of the program in the 2020–21 school year on the four outcomes of interest. The other eight schools were dropped from the Year 1 analysis because in the AP exam data available to the evaluation, the needed data for these eight schools were suppressed—either for the 2020 baseline measures (one school), the 2021 outcomes (six schools), or both (one school)—due to small cell sizes (specifically, having fewer than five exam takers).

Subsamples

We identified matched comparison schools separately within each of two subsamples of schools. The first subsample, termed Subsample 1, consists of schools that had AP CSP exam takers in the baseline year (2020), while Subsample 2 consists of schools that had did not have AP CSP exam takers in the baseline year (2020) but did have exam takers in *other* AP subjects that year. We conducted the matching for these two sets of schools separately—and, later, analyzed them separately—because they had different baseline measures. Subsample 1's baseline measures were based on 2020 AP CSP exam scores (or participation), while Subsample 2's baseline measures were based on scores (or participation) across *all* AP subject areas, as a proxy for AP CSP. Subsample 1 consists of a single block, while Subsample 2 consists of two blocks, as shall be discussed further below.

Matching Process

To match Lone Star schools with comparison schools that were equivalent at baseline on key measures (see below), we used the Stata module *kmatch* (Jann, 2017), which offers a wide array of matching methods and specifications. For all three matching blocks, we used multivariate-distance matching (option *md,* and specifically the default Mahalanobis distance matching) and specified nearest-neighbor matching (*nn*) without replacement (*wor*) to select a fixed number of unique, mutually exclusive comparison schools for each Lone Star school. The

[12] Respectively: https://nces.ed.gov/ccd/schoolsearch/, https://tealprod.tea.state.tx.us/Tea.AskTed.Web/Forms/SearchScreen.aspx?orgType=School, and https://apcourseaudit.inflexion.org/ledger/.

variables we matched on and the number of comparison schools we selected per each Lone Star school varied by block, as detailed in the following sections.

Subsample 1 Matching and Baseline Equivalence

Subsample 1's single matching block ("Block 1") was built around 13 Lone Star schools that had AP CSP exam takers *and* unsuppressed AP CSP exam data in both 2020 and 2021. They were matched with comparison schools of the same profile; there were 453 such schools in the five states for which we had data. Within this pool of potential comparison schools, we matched on the following four measures:

- Percentage of 2020 AP CSP exam takers earning a 3, 4, or 5 on the exam
- Percentage of students in grades 9–12 (in 2019–20) earning a 3, 4, or 5 on the 2020 AP CSP exam
- Percentage of students in grades 9–12 (in 2019–20) earning a 1 or 2 on the 2020 AP CSP exam
- Enrollment in grades 9 to 12 (in 2019–20)

The first three variables are baseline measures of the corresponding outcome variable. It was not necessary to explicitly match on the baseline measure of the fourth outcome—percentage of students in grades 9–12 taking the 2021 AP CSP exam—because this measure is, mathematically, the sum of the middle two measures, such that obtaining good matches on the middle two more or less ensured a good match on the fourth. (We did check this assumption after the matching process, and it held for all three blocks.[13]) For face validity, we also included enrollment in grades 9 to 12 (in 2019–20) as a matching variable, although we were prepared to employ less stringent criteria for baseline equivalence on this measure given that it was an ordinary covariate rather than a baseline measure of an outcome.

Because of the mathematical relationships among the outcome variables, we considered it essential to identify a single set of comparison schools to be used in the analyses for all four outcomes. Given this priority, we did not try to match on school demographic characteristics other than enrollment size; rather, we accounted for demographic differences between the two groups by including school demographic variables as covariates in our analysis models. The extent to which the Lone Star schools and the comparison schools were similar demographically is discussed later in this chapter.

[13] One might also think that matching on the middle two variables would also ensure a match on the first variable, which, at the individual school level, is the second variable divided by the sum of the second and third variables. However, because this relationship involves division, it does not hold up for *means* across multiple schools. That is, the mean of the first variable across schools does *not* equal the mean of the second variable divided by the sum of the second and third variable means.

We experimented with selecting 1 to 15 comparison schools as the number of comparison schools to be matched with each Subsample 1 Lone Star school and decided on 10 as the optimal number. Specifically, as selected by the *kmatch* command we used, the sample composed of 10 comparison schools matched with each Lone Star school offered excellent baseline equivalence on each of the four matching variables as well as considerable statistical power for the analysis. Thus, from the pool of 453 potential comparison schools, we selected 130 for the analysis.

Of the 130 selected comparison schools, 53 were in Texas.[14] Of the other 77, 13 were in Georgia, 23 were in Massachusetts, 25 were in New York, and 16 were in Wisconsin. Exhibit 5.1 displays the baseline means, standard deviations, and standardized mean differences (i.e., Hedges' *g*, as described in the box in Chapter 4) for the Subsample 1 Lone Star and selected comparison schools. The final column shows that for each of the four outcomes, the SMD on the baseline measure of the outcome was less than (an absolute value of) 0.05 standard deviations. Baseline differences of less than 0.05 standard deviations meet WWC criteria regardless of whether the baseline measure is included as a covariate in the impact analysis, but (as described below in the section titled "Statistical Estimation Strategy") we did nevertheless include the baseline measure as a covariate in the estimation model for its corresponding outcome. Exhibit D-1 in the appendix displays the equivalent information for schools' grades 9–12 enrollment sizes in the baseline year (not subject to WWC criteria, but on which we did nevertheless obtain strong baseline equivalence).

Exhibit 5.1. Baseline Statistics for the Year 1, Subsample 1 Analysis Sample

Measure	Lone Star schools (*n* = 13)	Comparison schools (*n* = 130)	*p* value	Hedges' *g*
Percentage of 2020 AP CSP exam takers earning a 3, 4, or 5 on the exam	65.30% (21.08)	65.30% (20.08)	1.000	0.000
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2020 AP CSP exam	1.19% (0.90)	1.23% (1.04)	0.893	-0.039
Percentage of students in grades 9–12 earning a 1 or 2 on the 2020 AP CSP exam	0.79% (1.38)	0.76% (1.26)	0.939	0.022
Percentage of students in grades 9–12 taking the 2020 AP CSP exam	1.98% (1.85)	1.99% (1.90)	0.983	-0.006

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from independent-samples *t* tests; they are provided for parallelism with the outcome analysis results. Hedges' *g* in the last column is calculated using the formula provided in Chapter 4, except that the *N*s are numbers of schools rather than numbers of teachers.

[14] Given that all of the Lone Star schools were in Texas, it was essential that at least some of the selected comparison schools also be in Texas.

Subsample 2 Matching and Baseline Equivalence

As noted earlier, the schools in Subsample 2 did not have AP CSP exam takers in 2020 but had takers of *other* subjects' AP exams in 2020, with unsuppressed data on the total number of AP exams taken (across all AP subjects) and score distributions. Subsample 2 consists of two blocks of schools. In both blocks, Lone Star schools were matched with comparison schools that, similarly, did *not* have AP CSP exam takers in 2020, and also similarly, had unsuppressed 2020 data (number of exams taken and score distributions) on all AP exams.

The two blocks in Subsample 2 differ based on the Lone Star schools' *2021* profiles. Block 2a was based on 11 Subsample 2 Lone Star schools that had AP CSP exam takers, with unsuppressed data, in 2021; they were matched with comparison schools of the same profile, of which there were 91 across the five states. Block 2b was based on four Subsample 2 Lone Star schools that did not have AP CSP exam takers in 2021 (in addition to not having had AP CSP exam takers in 2020). They were matched with comparison schools that did not have *suppressed* AP CSP data in 2021; that is, the potential comparison schools either had *no* AP CSP exam takers in 2021 (like the Block 2b Lone Star schools) *or, unlike* the Block 2b Lone Star schools, had unsuppressed 2021 AP CSP exam data.[15] Because unselected schools in the Block 2a pool met the criteria to be considered as matches in Block 2b, we conducted the matching for the two blocks sequentially, with Block 2a first. Potential comparison schools that were not selected for Block 2a were added to the matching pool for Block 2b, which then consisted of 1,528 schools.

For both Subsample 2 blocks, we matched on their average 2020 AP exam performance across all AP subject areas,[16] corresponding to the AP CSP outcomes, as proxy measures (see the box titled "On Proxy Measures and Outcome Domains"):

---

[15] If the comparison schools in this block had been restricted to those that, like the Lone Star schools in the block, had no AP CSP exam takers in 2021, both the Lone Star schools and the comparison schools would have had zeroes for each of the four outcome measures. Because the impact analysis includes block indicators, Lone Star schools and comparison schools are effectively compared only within block, so there would have been no point in intentionally including a block of Lone Star schools and comparison schools known *a priori* to be identical on all outcomes. In addition, a key goal of the Lone Star AP CSP program was to increase the number of students having access to AP CSP, so it made sense to allow comparison schools that had AP CSP exam takers in the outcome year even knowing that some Lone Star schools did not. In forming the pool of potential comparison schools for Block 2b, we eliminated all schools that had *suppressed* 2021 AP CSP exam score data, but we purposely remained blind to whether or not the remaining schools had AP CSP exam takers in 2021.

[16] Initially we had planned to calculate the average across only the AP subject areas that had exam takers in any given school, but the Texas data we received from the College Board did not break out specific AP subjects other than AP CSP. Moreover, in processing the data from the other states, which did break out each individual AP subject area, we noticed that it was quite common for schools to have just one or two exam takers for some subject areas (but more for other subject areas), suggesting that perhaps schools may have had exam takers in AP subjects that were not actually taught at the school. Because we had no source of data on what AP subject area courses were *offered* by each school, we decided instead to average the exam outcomes across all 38 AP subject areas for each school. This decision had the added advantage that it enabled us to calculate the cross-subject averages from the AP exam total counts (across all subject areas), which were available in the data from all five states and which, because the total counts were larger than the counts for individual subject areas, were less likely to be suppressed.

- Average percentage (across all 38 AP subject area exams) of 2020 exam takers earning a 3, 4, or 5
- Average percentage (across all 38 AP subject area exams) of students in grades 9–12 (in 2019–20) earning a 3, 4, or 5 in 2020
- Average percentage (across all 38 AP subject area exams) of students in grades 9–12 (in 2019–20) earning a 1 or 2 in 2020

As in Subsample 1, it was not necessary to explicitly match on the average percentage of students *taking* a 2020 AP exam, given that a match on this measure was more or less ensured by matches on the second and third outcomes. Also as in Subsample 1, baseline-year enrollment size in grades 9 through 12 was included as an additional matching variable in the Subsample 2 matching process, but other school demographic characteristics were not. (School demographics are discussed later in this chapter.)

ON PROXY MEASURES AND OUTCOME DOMAINS

WWC (2022) states: "If the outcome is a broad measure of knowledge or skills, such as standardized academic achievement measures, then using a broad baseline measure of knowledge or skill is sufficient for satisfying the baseline equivalence standard, and therefore authors *would not need to satisfy the standard for additional student or contextual characteristics, even if the baseline measure is in a different outcome domain*" (p. 55, emphasis added). Because the AP CSP exam is a standardized academic achievement measure, the corresponding Subsample 2 performance measure across *all* AP exams—i.e., "a broad baseline measure of knowledge or skill"—can, on its own, satisfy the baseline equivalence standard for each of the three performance outcomes in the Subsample 2 analyses.

A note on domains: We consider the AP CSP performance outcomes to be in the "Technology and Engineering" domain of the "Academic Readiness, Knowledge, or Skills (Preschool Through Postsecondary)" domain category, while the Subsample 2 "all AP exams" baseline performance measures fall into the "General Achievement: Academic Achievement PK–12" domain of the same larger category (WWC, 2024). The participation outcome (AP CSP exam taking) and the corresponding Subsample 2 baseline measure (any AP exam taking) are both in the same domain—"Access to Educational Opportunities," within the School and Family Outcomes and Educational Opportunity category (WWC, 2024)—and therefore AP exam taking across all subject areas can unquestionably constitute the sole baseline measure for the AP CSP exam taking outcome in the Subsample 2 analyses.

For Block 2a, we experimented with selecting up to seven comparison schools as the number of comparison schools to be matched with each Lone Star school and decided on three as the optimal number (balancing strong baseline equivalence results with statistical power), resulting in a total of 33 comparison schools, 25 of which were in Texas. For Block 2b, we experimented with selecting up to 10 comparison schools per Lone Star school and settled on six as a ratio offering good baseline equivalence and statistical power, resulting in a total of 24 comparison schools, 14 of which were in Texas.[17] Exhibit 5.2 shows the states of the selected comparison schools in each block.

Exhibit 5.2. States of the Year 1, Subsample 2 Comparison Schools

Block	Texas	Georgia	Massachusetts	New York	Wisconsin	Total
2a	25	1	3	2	2	**33**
2b	14	3	2	2	3	**24**
Subsample 2 Total	**39**	**4**	**5**	**4**	**5**	**57**

The within-block baseline equivalence statistics for each of the Subsample 2 blocks, which were used to select the comparison schools within each block, are provided in Appendix Exhibit D-2. Because the two blocks are combined in the Subsample 2 impact analyses (with a block indicator, as will be described in the estimation strategy section), we focus here on demonstrating baseline equivalence for the subsample as a whole.

To calculate the combined-block baseline equivalence, we conducted a linear regression for each of the baseline measures, where the dependent variable was the value of the baseline measure, the key predictor was an indicator for Lone Star schools, and a key covariate was a dummy variable for matching block. The equation for the regression was as follows:

$$Y_s = \beta_0 + \beta_{Trt} Lonestar_s + + \beta_{Block} Block_s + \varepsilon_s$$

where Y_s represents the baseline measure of interest in school *s*, $Lonestar_s$ is a binary variable indicating whether school s is a Lone Star school, $Block_s$ is a binary variable indicating whether school s was in matching Block 2b (as opposed to Block 2a) and ε_s is the error term. The coefficient β_{Trt} is the block-adjusted mean difference between the Lone Star and comparison groups on the baseline measure.

[17] As noted in an earlier footnote, we purposely remained blind, during the matching, to whether the schools in the Block 2b comparison pool had AP CSP exam takers in 2021. It turned out that of the 24 selected schools, none of them had AP CSP exam takers in 2021.

The Subsample 2 combined-block baseline equivalence results are presented in Exhibit 5.3. As shown in the last column, the Hedges' *g* for each baseline measure was less than (an absolute value of) 0.25 standard deviations, which meets WWC criteria provided that the baseline measure is included as a covariate in the impact analysis for its corresponding outcome measure, which it was (see "Statistical Estimation Strategy," below).

Exhibit 5.3. Baseline Statistics for the Year 1, Subsample 2 Analysis Sample

Measure	Lone Star schools (*n* = 15)	Comparison schools (*n* = 57)	Regression coefficient (β_{Trt})	*p* value	Hedges' *g*
Average percentage (across all 38 AP subject area exams) of 2020 exam takers earning a 3, 4, or 5	36.57% (15.69)	37.10% (16.52)	-0.93%	0.846	-0.057
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 3, 4, or 5 in 2020	0.31% (0.26)	0.32% (0.31)	-0.03%	0.740	-0.096
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 1 or 2 in 2020	0.49% (0.39)	0.46% (0.33)	-0.00%	0.960	-0.014
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 taking a 2020 AP exam	0.80% (0.55)	0.79% (0.58)	-0.03%	0.832	-0.060

Note. In the first two columns, each cell contains the unadjusted mean and (in parentheses) standard deviation of the baseline measure. In the third column, the regression coefficient is the coefficient on the Lone Star indicator from the regression of the baseline measure on the Lone Star indicator and matching block indicator (i.e., the block-adjusted baseline mean difference). The Hedges' *g* value results from the inclusion of the regression coefficient as the numerator (i.e., the block-adjusted mean difference) in the Hedges' *g* formula.

In summary, for both of the two subsamples, we succeeded in selecting a well-matched set of comparison schools—equivalent to the Lone Star schools at baseline on the outcomes to be examined—for the 2021 analysis of program impacts on AP CSP exam performance and participation.

Statistical Estimation Strategy

To estimate the effect of the Lone Star program on the 2021 AP CSP exam performance and participation outcomes (i.e., impact after one year of program implementation), we began by running a separate ordinary least squares (OLS) regression for each of the four outcomes in

each of the two subsamples.[18] As discussed earlier, the unit of analysis was the school. For Subsample 1, the analytic model had the following general specification:

$$y_s = \beta_0 + \beta_{Trt} Lonestar_s + \beta_{Pre} PreOutcome_s + \boldsymbol{x_s'\beta_{Cov}} + \varepsilon_s$$

where y_s represents the outcome of interest of school *s*, $Lonestar_s$ is a binary variable indicating whether school s is a Lone Star school, and $PreOutcome_s$ represents the 2020 (i.e., pre-intervention, or baseline) measure of the outcome. The vector $\boldsymbol{x_s'}$ represents school-level characteristics (see below), and ε_s is the error term. The coefficient β_{Trt} represents the estimated intervention effect.

The analytic model for Subsample 2 was similar, the only differences being that:

a. It contained one additional term: a binary variable indicating whether school s was in matching Block 2b (as opposed to Block 2a);

b. The baseline measure was the proxy measure of the 2020 average across all 38 AP exams, because Subsample 2 schools did not have AP CSP exam takers in 2020.

After conducting these regressions, we calculated the effect size of the estimated intervention effect for each of the four measures in each of the two subsamples. Again we used the formula for Hedges' *g* as described in Chapter 4, except that instead of the numerator being the *un*adjusted mean difference between the Lone Star and comparison groups at baseline, it was the *adjusted* mean difference in the *outcome* year, namely the β_{Trt} coefficient from the regression. (The standard deviations in the denominator were also from the outcome year.)

Covariates

All models included the same set of school characteristic covariates, decided *a priori:*

[18] We had initially planned to include both subsamples in a single analysis by standardizing the baseline measure within block and including block indicators (indicating Subsample 1, Block 2a, and Block 2b) in the regression model. However, we decided against this plan because although the block indicators would allow the intercept to vary by block, there would still be a single slope coefficient for the baseline measure. In other words, the relationship between the baseline measure and the outcome variable would be constrained to be the same across blocks. We did not think that simply standardizing the baseline measure (within block) was sufficient to warrant the assumption that the AP CSP baseline measure used for Subsample 1 would have the same relationship to the corresponding AP CSP outcome as the all-AP-exams baseline measure used for Subsample 2. We could have used interaction terms to allow the slope of the baseline measure to vary by block but decided that separate analyses would be a simpler solution. Moreover, the separation gives reviewers, including WWC reviewers, the flexibility to conduct separate reviews for each of the two subsamples.

- Total student enrollment in the 2020–21[19] school year
- Percentage of the 2020–21 school enrollment that was Black
- Percentage of the 2020–21 school enrollment that was Hispanic
- Percentage of the 2020–21 school enrollment that was female
- Within-state standardized percentage of the 2020–21 school enrollment that was economically disadvantaged[20]
- Whether the school was a charter school
- Locale type: suburb, town/rural, or (reference group) city
- An indicator of missingness for the economically disadvantaged covariate[21] (the only covariate that had any missing data)

All of these data were drawn from the Common Core of Data, except for the economically disadvantaged variable for Massachusetts schools, which came from state data.[22] For the percentage variables, we calculated the percentages based on counts provided in the administrative data.

Appendix Exhibits D-3 and D-4 provide descriptive statistics on all of the covariate measures for the Subsample 1 Lone Star and comparison schools, and Appendix Exhibits D-5 and D-6 provide the covariate measure descriptives for the Lone Star and comparison schools in Subsample 2. As evidenced by the effect sizes in the rightmost column of each exhibit, there were some demographic measures on which the Lone Star schools and comparison schools were not very similar. In particular, in both Subsamples 1 and 2, Lone Star schools were much more likely than comparison schools to be charter schools. In Subsample 1, Lone Star schools, on average, had

[19] We used the outcome year (in this chapter, 2020–21) rather than the 2019–20 baseline year as the year for all school demographic characteristics, because it makes little conceptual sense to control for the demographics of a different population of students. For example, if each grade level in a school with grades 9–12 had approximately the same number of students and all students moved up each year, the school population would be only 75% the same between the baseline year and the first outcome year, and only 25% the same between the baseline year and the third outcome year. Unlike with AP CSP classroom composition as discussed in Chapter 4, nothing in the Lone Star AP CSP program aims to affect *school* composition, and it is highly unlikely that a program targeting a single AP course would have an *unintentional* effect on school composition. Thus, there is no strong methodological argument against using the outcome-year school demographics.

[20] For schools in all states except Massachusetts, the economically disadvantaged covariate was the school percentage of students in the free and reduced-price lunch (FRL) program, based on counts provided in the Common Core of Data (CCD). However, FRL data were not available in CCD for Massachusetts schools. For Massachusetts schools, we instead used percentages of "economically disadvantaged" students, based on counts in data files at https://profiles.doe.mass.edu/statereport/selectedpopulations.aspx. Because of the difference between Massachusetts and the other states on this variable, we standardized the values within state.

[21] One New York school was missing economically disadvantaged data; we imputed the value as zero (i.e., the within-state standardized mean). This school was in Subsample 1, so the missingness indicator dropped out of the Subsample 2 analyses.

[22] Common Core of Data: https://nces.ed.gov/ccd/elsi/tableGenerator.aspx. Massachusetts economically disadvantaged: https://profiles.doe.mass.edu/statereport/selectedpopulations.aspx

much higher proportions of Hispanic students, slightly lower proportions of female students, and slightly higher percentages of economically disadvantaged students than comparison schools. They were also notably more likely to be located in cities and less likely to be located in suburbs. In Subsample 2, Lone Star schools had slightly higher percentages of Black students and lower percentages of female students, on average, than comparison schools; they were also slightly more likely to be located in cities, notably more likely to be in suburbs, and much less likely to be in rural areas or towns. By including school demographic characteristics as covariates in the statistical models, the analyses adjusted for any differences between Lone Star schools and comparison schools on these characteristics.

Meta-Analysis

Although we conducted separate regression analyses for the two subsamples of schools, our ultimate aim was to describe the impact of the Lone Star AP CSP program on the entire set of Lone Star schools as a single group, regardless of whether they had AP CSP in the baseline year. Thus, for each outcome measure, we conducted a meta-analysis combining the estimates from the two subsamples to yield a single overall estimate. The input for the meta-analysis of each outcome consisted of the effect size (Hedges' *g*), as calculated for each of the two subsamples based on the OLS regression results, along with the standard error of the effect size calculated in the same manner. We used the Stata command *meta* to conduct each meta-analysis and output forest plots to display the results.[23]

Results

In this section we first present the results separately for each of the two subsamples and then present the meta-analysis (combined) results.

Subsample 1 and 2 (OLS-Based) Results

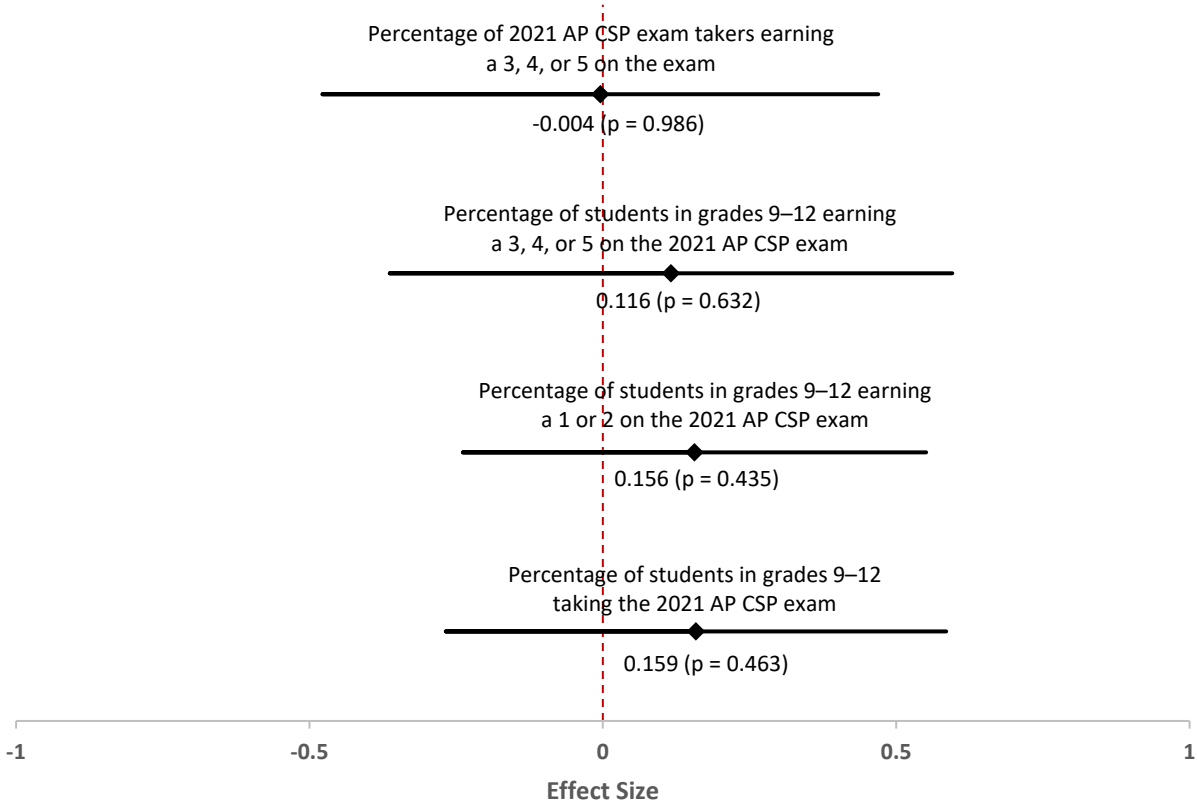
Exhibits 5.4 and 5.5 show the estimated effect sizes for the first-year effects of the Lone Star AP CSP program on the four outcome measures for Subsample 1 and Subsample 2, respectively. Exhibit 5.6 provides additional details, including the unadjusted (i.e., descriptive) means and standard deviations of each outcome measure, the regression-adjusted mean difference (i.e., the coefficient on the Lone Star indicator from the OLS regression), the *p* value of the coefficient from the regression, and the calculated Hedges' *g* effect size (identical to the point estimates shown in Exhibits 5.4 and 5.5). The full regression results are provided in the appendix (Exhibit D-7 for Subsample 1 and Exhibit D-8 for Subsample 2).

As shown in Exhibit 5.4, after being implemented for one year, the Lone Star program did not have a significant positive effect on any of the four outcome measures for the schools in

[23] We used Stata's default settings for the meta-analysis, e.g., random effects with restricted maximum likelihood estimation.

Subsample 1 (i.e., those that had AP CSP exam takers in the baseline year). The effect was essentially zero for the key performance outcome, the percentage of AP CSP exam takers earning a qualifying score on the exam. For the two supplementary performance measures, the observed effects were positive but relatively small. (Recall that for the second of the two supplementary performance measures, the percentage of students in grades 9–12 earning a score of 1 or 2, the effect would be theorized to be *negative* if the program was successful). The same was true for the participation outcome.

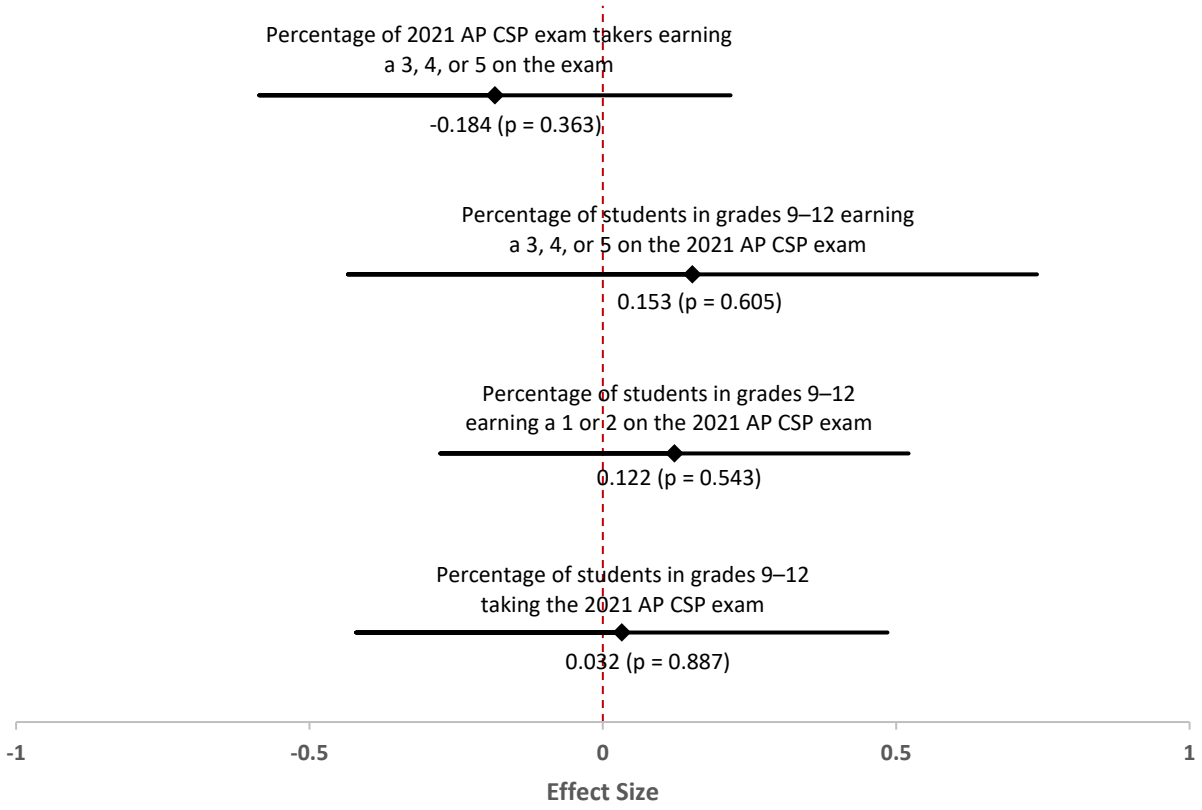
Exhibit 5.4. Effect Size Chart for the Year 1, Subsample 1 Analysis



Note. *N* = 143 (13 Lone Star schools and 130 comparison schools). Effect size point estimates are represented by the diamonds; the horizontal lines show 95% confidence intervals. Confidence intervals that do not intersect with an effect size of 0 (represented by the dotted red line) indicate a statistically significant ($p < .05$) effect.

Exhibit 5.5 shows that there were also no significant effects for the Subsample 2 schools (i.e., those that did not have AP CSP exam takers in the baseline year). The effect for the key performance measure—the percentage of AP CSP exam takers earning a qualifying score on the exam—was negative but relatively small. The effects for the two supplementary performance measures were, as in Subsample 1, positive but small. The effect for the participation outcome was close to zero.

Exhibit 5.5. Effect Size Chart for the Year 1, Subsample 2 Analysis



Note. *N* = 72 (15 Lone Star schools and 57 comparison schools). Effect size point estimates are represented by the diamonds; the horizontal lines show 95% confidence intervals. Confidence intervals that do not intersect with an effect size of 0 (represented by the dotted red line) indicate a statistically significant ($p < .05$) effect.

Exhibit 5.6. Year 1 Outcome Statistics for Both Subsamples

Measure	Lone Star schools (unadj.)	Comparison schools (unadj.)	Adj. mean difference (β_{Trt})	*p* value	Hedges' *g*
Subsample 1 (13 Lone Star schools and 130 comparison schools)					
Percentage of 2021 AP CSP exam takers earning a 3, 4, or 5 on the exam	52.36% (21.49)	64.84% (26.01)	-0.11% (6.17)	0.986	-0.004
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2021 AP CSP exam	1.28% (1.33)	1.29% (1.28)	0.15% (0.31)	0.632	0.116
Percentage of students in grades 9–12 earning a 1 or 2 on the 2021 AP CSP exam	1.63% (3.12)	0.87% (1.63)	0.28% (0.36)	0.435	0.156
Percentage of students in grades 9–12 taking the 2021 AP CSP exam	2.91% (4.32)	2.15% (2.34)	0.41% (0.56)	0.463	0.159
Subsample 2 (15 Lone Star schools and 57 comparison schools)					
Percentage of 2021 AP CSP exam takers earning a 3, 4, or 5 on the exam	24.30% (26.02)	24.65% 30.46%	-5.51% (6.01)	0.363	-0.184
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2021 AP CSP exam	1.28% (2.47)	0.56% 1.21%	0.24% (0.46)	0.605	0.153
Percentage of students in grades 9–12 earning a 1 or 2 on the 2021 AP CSP exam	3.71% (10.67)	1.20% 3.59%	0.71% (1.16)	0.543	0.122
Percentage of students in grades 9–12 taking the 2021 AP CSP exam	4.99% (11.99)	1.77% 4.36%	0.22% (1.52)	0.887	0.032

Note. In the first two columns, each cell contains the unadjusted mean and (in parentheses) standard deviation of the outcome measure. The third column contains the regression coefficient on the Lone Star indicator from the impact analysis model (i.e., the adjusted mean difference) and (in parentheses) the standard error of the coefficient; the *p* value associated with the coefficient is in the fourth column. The Hedges' *g* value in the final column results from the inclusion of the regression coefficient as the numerator (i.e., the mean difference) in the Hedges' *g* formula.

As shown in Appendix Exhibits D-7 and D-8, the baseline measure of the outcome was a significant predictor of all four outcome measures in both subsamples. None of the school demographic covariates significantly predicted all four outcomes in either subsample, but there were some that significantly predicted one to three outcomes in one subsample or the other.

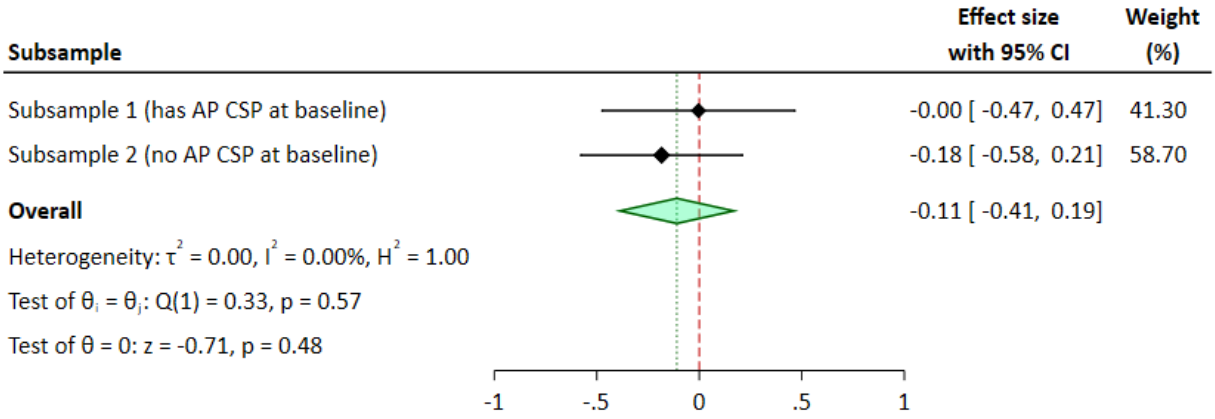
Across the two subsamples, there are no clear patterns of significant relationships between particular school characteristics and any of the outcomes. This finding lends some support to our decision not to include these characteristics in our matching process for identifying comparison schools, but more analyses are needed to conclude that the characteristics are not generally predictive of any of the outcomes (at least when the baseline measure of the outcome is also included as a covariate). Accordingly, subsequent years' analyses (see Chapters 6 and 7) continue to include the school characteristics as covariates.

Meta-Analysis Results

Exhibits 5.7 through 5.10 display the Year 1 meta-analysis results for each of the four outcome measures, respectively. Not too surprisingly given the nonsignificant and small effects found for the two individual subsamples, the meta-analysis, too, showed no significant impact for any of the four outcome measures after one year of program implementation, and all four overall effect size estimates are small (ranging from -0.11 to 0.14). The estimate for the key performance outcome—the percentage of AP CSP exam takers earning a qualifying score on the exam—was negative, and the estimate for the second supplementary performance outcome—the percentage of students in grades 9–12 earning a score of 1 or 2—was positive; both of these are in the opposite direction from the hypothesized effect if the program was successful. That said, the lack of statistical significance means that we cannot, with confidence, rule out that the true effect (as opposed to the estimated effect) was actually null, or zero. The same is true for the other two outcomes—the first supplementary performance outcome and the participation outcome—for which the observed effects are in the hypothesized direction (positive for both).

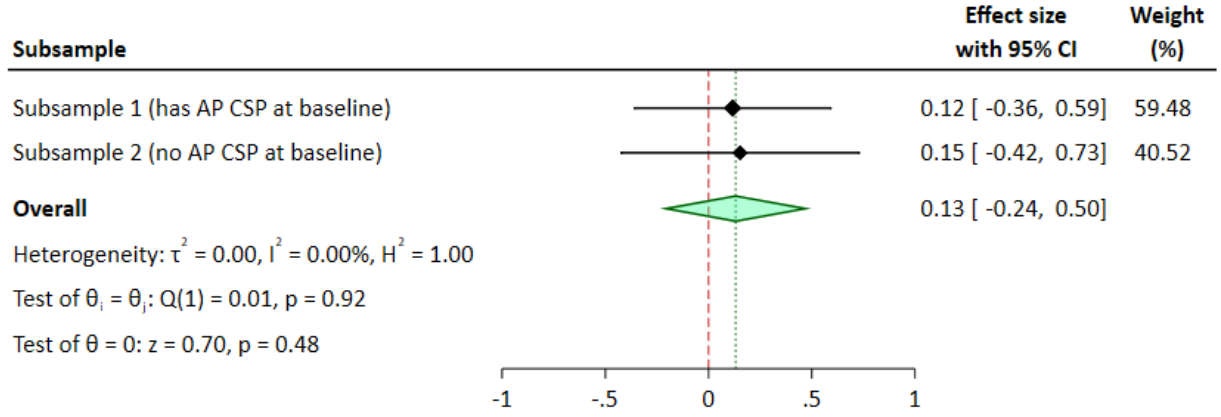
The meta-analysis results also indicate that there was no significant variability across the two subsamples for any of the four outcomes, and in no case was the effect for either subsample significantly different than the overall effect.

Exhibit 5.7. Meta-Analysis Plot for Percentage of 2021 AP CSP Exam Takers Earning a 3, 4, or 5 on the Exam



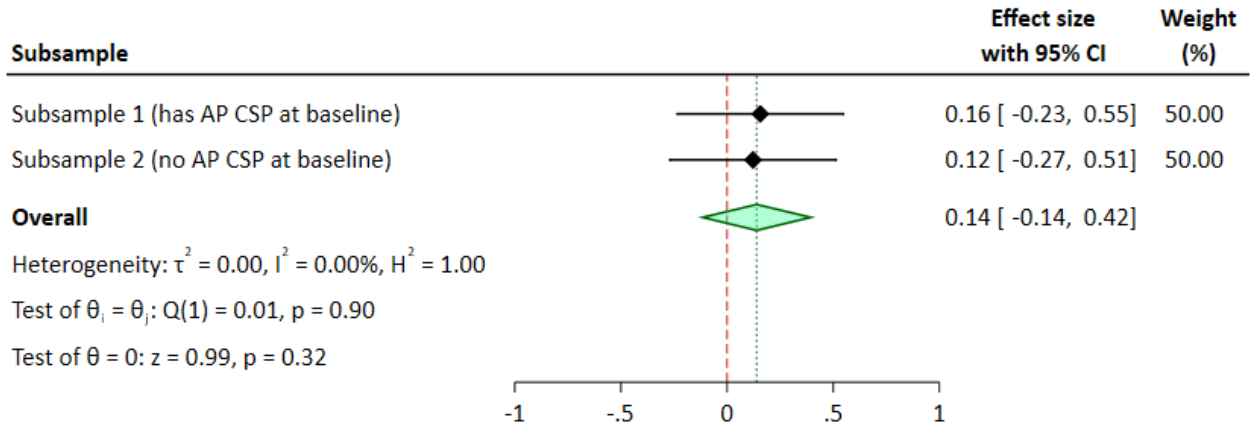
Note. See box titled "How to Read the Meta-Analysis Figures."

Exhibit 5.8. Meta-Analysis Plot for Percentage of Students in Grades 9–12 Earning a 3, 4, or 5 on the 2021 AP CSP Exam



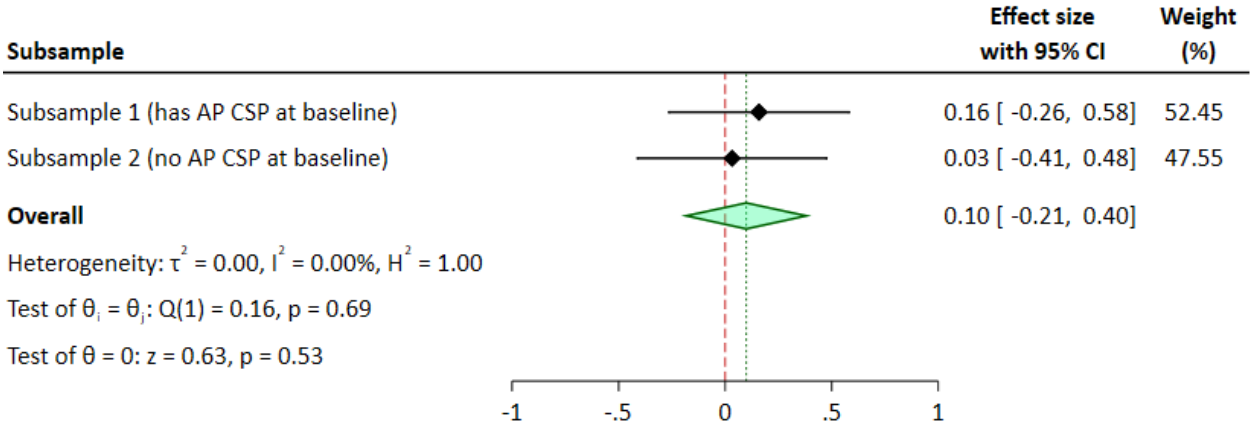
Note. See box titled "How to Read the Meta-Analysis Figures."

Exhibit 5.9. Meta-Analysis Plot for Percentage of Students in Grades 9–12 Earning a 1 or 2 on the 2021 AP CSP Exam



Note. See box titled "How to Read the Meta-Analysis Figures."

Exhibit 5.10. Meta-Analysis Plot for Percentage of Students in Grades 9–12 Taking the 2021 AP CSP Exam



Note. See box titled "How to Read the Meta-Analysis Figures."

HOW TO READ THE META-ANALYSIS FIGURES

Figures 5.7 through 5.10 were created using the Stata command *meta forestplot.* The following is a guide to how to read them:

- The black lines, black diamonds, and subsample effect size point estimates are essentially the same as those shown in Exhibits 5.4 and 5.5 for the corresponding outcome and subsample. For all of the figures, we have set the horizontal axis range to run from −1 to +1 effect size units (i.e., standard deviations of the outcome variable). This range was sufficient to capture the full width of each CI.
- For each subsample, the number in the "Weight (%)" column on the far right is the percentage of the total weight that the subsample was given in the meta-analysis. Each subsample's weight is a function of the precision of the subsample's estimate; the smaller the standard error (and CI), the larger the weight. The relative weights of the subsamples are also reflected in the size of the black-diamond point estimates, with larger weights reflected by larger diamonds. If the two subsamples have equal weight, their black diamonds are the same size.
- The vertical green small-dotted line shows the point estimate for the overall effect size calculated by the meta-analysis (θ), and the horizontal width of the large green diamond is the 95% CI for the meta-analysis point estimate. (The numeric values of the point estimate and the CI endpoints are shown in the "Effect size with 95% CI" column.) If the diamond does not intersect with the vertical dashed red line at zero, the overall effect is significant at $p < .05$. The p value of the significance test is shown in the "Test of θ = 0" under the "Overall" heading.
- If the black horizontal line for a subsample does not cross the vertical green small-dotted line, the effect for that subsample is significantly different (at $p < .05$) from the overall effect.
- The heterogeneity statistics under the "Overall" heading indicate how much variability (or heterogeneity) there is across the subsamples, with the I^2 percentage indicating the percentage of the variability that is due to a true difference between the subsample effects rather than random sampling error. The test of $\theta_i = \theta_j$ (also under the "Overall" heading) indicates whether the variability across subsamples (if any) is statistically significant.

Summary and Discussion

In summary, we found that after one year of implementation, the Lone Star AP CSP program had no statistically significant effect on any of the three performance outcomes or on the participation outcome. This was true for the overall sample of schools included in the analysis—28 Lone Star AP CSP schools and 187 matched comparison schools that were similar to the Lone Star schools at baseline on each outcome—as well as individually for each of the two subsamples of schools that constituted the overall sample.

Findings described in the prior chapters of this report may offer some insights into the lack of significant positive effects. For the performance outcomes, which were theorized to be improved indirectly by the Lone Star AP CSP program's supports for teachers and directly by the supports for students (see Exhibit 2.1), the implementation study described in Chapter 2 found that the teacher supports component of the program was not implemented with fidelity in Year 1. The student supports component was implemented with fidelity that year, but the fidelity

measures for the student supports component mainly gauged whether the student supports were offered rather than the extent to which they were actually used. As discussed in Chapter 3, teacher participation in some program activities and perceptions of the usefulness of the activities were both relatively low in Year 1, as was uptake of the student resources, according to Lone Star AP CSP teachers who responded to a survey administered at the end of that year. And in Chapter 4 we found that in the first year of implementation, the program had a significant effect on only one of three measures of teachers' instructional practices in their AP CSP courses. Taken together, these findings likely help explain the null effects on student performance outcomes in the program's first year.

We have less context to explain the null finding for the participation outcome in Year 1. According to the logic model, the program's supports for schools were theorized to increase the percentages of students taking the AP CSP course and exam. As described in Chapter 2, the evaluation found that the school supports component was implemented with fidelity in Year 1, but the specific elements of the component that met the fidelity criteria were those that examined provision of resources to the schools. We did not have data on how the resources were used or on levels of program participation or uptake by school staff. That said, students most likely enrolled in AP CSP at the beginning of the school year (or even at the end of the previous school year), before most of the Year 1 supports for *any* of the three program components were even put into play, so this could help explain why there was no significant effect on participation, especially insofar as students who took the course also took the exam.

As has already been discussed in prior chapters, the first year of the program was the 2020–21 school year, which was severely disrupted by the COVID-19 pandemic. The pandemic-caused disruptions and upheaval no doubt made it challenging for many new academic programs, especially those not directly focused on mitigating pandemic effects, to have much of an impact on student outcomes that year.

Moreover, given the program's emphasis on a full three-year arc of professional development for teachers, significant positive effects after just a single year may not even have been expected, realistically speaking. Larger effects would have been theorized to occur in the following years; the extent to which these larger effects were realized is discussed in the next two chapters.

Chapter 6. Impact on Student AP CSP Performance and Participation After Two Years

This chapter examines whether the program achieved its aims for AP CSP performance and participation (see Exhibit 2.1 in Chapter 2) during its second year of operation, the 2021–22 school year. The research question addressed is as follows:

5. What was the impact of two years of participation in the Lone Star AP CSP program, compared to business-as-usual AP CSP participation, on high schools' rates of (a) AP CSP exam takers receiving a qualifying score on the AP CSP exam, and (b) students taking the AP CSP exam?

The preceding chapter looked at the impact after one year of implementation, and the following chapter will look at the impact after three years of implementation.

As described in the preceding chapter, the student impact study used a quasi-experimental matched comparison group design and an analysis-of-covariance (regression) estimation model to examine the effect of Lone Star AP CSP on the performance and participation outcomes, followed by meta-analysis to combine the estimates from two samples of schools (those that had AP CSP exam takers in the baseline year and those that did not).

KEY FINDINGS

- After two years of implementation, the Lone Star AP CSP program had a borderline-significant negative effect on schools' percentages of AP CSP exam takers that received a qualifying score on the exam.
- The program had a significant negative effect on schools' percentages of students in grades 9–12 that took the AP CSP exam in the second year of implementation.
- Likely because of the negative effect on the percentages of students taking the exam, the program also had a significant negative effect on schools' percentages of students that received a qualifying score on the AP CSP exam, as well as a borderline-significant negative effect on schools' percentages of students that received a *non*qualifying score.

The remainder of this chapter is organized in the same way as the preceding chapter. First, we recap the measures of AP CSP achievement and participation we examined as outcomes, along with our sources of data on these measures, although we do not go into as much depth as when we first introduced the measures in Chapter 5. We then discuss the sample of Lone Star schools included in the analysis and explain how we identified comparison schools that were similar at baseline to the Lone Star schools. We then recap our

statistical estimation strategy, followed by presentation of the results. The chapter concludes with a brief summary and discussion.

Outcome Measures

The outcome measures examined in the analysis of the impacts of the second year of Lone Star AP CSP—the 2021–22 school year—are the same as those described in Chapter 5, except that they are based on the AP CSP exam administered in spring 2022 rather than in spring 2021:[1]

- the percentage of AP CSP exam takers earning a 3, 4, or 5 on the exam *(key measure of student performance)*;
- the percentage of students in grades 9–12 earning a 3, 4, or 5 on the AP CSP exam *(supplementary measure #1 of student performance)*;
- the percentage of students in grades 9–12 earning a 1 or 2 on the AP CSP exam *(supplementary measure #2 of student performance)*; and
- the percentage of students in grades 9–12 taking the AP CSP exam *(measure of student participation)*.[2]

The data sources for the outcome measures and their corresponding baseline measures were also the same as in the first-year impact analysis:

- The AP data for schools in Texas (including the Lone Star schools) came from datasets prepared by the College Board.
- The AP data for schools in four other states—Georgia, Massachusetts, Wisconsin, and New York—came from datasets posted publicly on those states' websites.[3]
- Enrollment data, used to calculate the percentages for the participation measure and the two supplementary performance measures, came from the websites of the five states.

See Chapter 5 for further detail.

Lone Star School Sample and Selection of Comparison Schools

As noted in earlier chapters, 37 Texas schools were recruited by NMSI to participate in the Lone Star AP CSP intervention, and 36 of them were eligible for inclusion in the AP exam impact analyses. Of these 36 Lone Star schools, 34 were included in the analyses of the impact of the

[1] For the three measures using grades 9–12 enrollment as the denominator for the percentage calculations, school enrollments were from the same school year (2021–22).
[2] See Chapter 5 for further detail about the four measures, including discussion of how they are mathematically related to one another and why this is potentially important in the interpretation of results.
[3] See footnote 6 in Chapter 5 for further explanation of and rationale for why we allowed some comparison schools to be located outside of Texas.

program in the 2021–22 school year on the four outcomes of interest. The other two schools, both of which had AP CSP exam takers in the 2019–20 (baseline) year, could not be included in the Year 2 analysis because their 2020 AP CSP scores were suppressed due to small cell sizes (i.e., fewer than five exam takers). It is worth noting that in the Year 1 analyses (see Chapter 5), only 26 Lone Star schools could be included, so the 34 schools in the two-year analyses is a notable increase.

Subsamples and Matching Process

As in the analyses of impact after one year, the identification of matched comparison schools, as well as much of the subsequent analysis, occurred separately within each of two subsamples of schools:

- Subsample 1 consists of schools that had AP CSP exam takers in the baseline year (2020).
- Subsample 2 consists of schools that had did not have AP CSP exam takers in the baseline year (2020) but did have exam takers in *other* AP subjects that year.

We again used the Stata module *kmatch* (Jann, 2017) to match Lone Star schools with comparison schools that were equivalent at baseline on key measures. See Chapter 5 for further detail.

Subsample 1 Matching and Baseline Equivalence

Subsample 1's single matching block ("Block 1") was built around 13 Lone Star schools that had AP CSP exam takers *and* unsuppressed AP CSP exam data in both 2020 and 2022.[4] They were matched with comparison schools of the same profile; there were 472 such schools in the five states for which we had data. Within this pool of potential comparison schools, we matched on the same four measures as in the Year 1 analysis (see Chapter 5) and, as in Year 1, considered it essential to identify a single set of comparison schools to be used in the analyses for all four outcomes.

We experimented with selecting 1 to 15 comparison schools as the number of comparison schools to be matched with each Subsample 1 Lone Star school and decided on a 1:7 intervention-to-comparison ratio to provide both strong baseline equivalence on each of the four matching variables and statistical power for the analysis. Thus, from the pool of 472 potential comparison schools, we selected 91 for the analysis.

[4] These 13 schools were the same 13 schools that were in Subsample 1 of the Year 1 analysis (Chapter 5). We decided to rematch—that is, identify a new set of comparison schools—because some comparison schools selected as matches in the analysis of the one-year impact may not have had AP CSP exam takers in the second year (i.e., the 2021–22 school year).

Of the 91 selected comparison schools, 34 were in Texas, 8 were in Georgia, 19 were in Massachusetts, 17 were in New York, and 13 were in Wisconsin. Exhibit 6.1 displays the baseline means, standard deviations, and standardized mean differences (i.e., Hedges' *g*) for the Subsample 1 Lone Star and selected comparison schools. The final column shows that for each of the four outcomes, the SMD on the baseline measure of the outcome was less than (an absolute value of) 0.05 standard deviations. As has been previously noted, baseline differences of less than 0.05 standard deviations meet WWC criteria regardless of whether the baseline measure is included as a covariate in the impact analysis, but we nevertheless did include the baseline measure as a covariate in the estimation model for its corresponding outcome (see "Statistical Estimation Strategy" section below). Exhibit E-1 in the appendix displays the equivalent information for schools' grades 9–12 enrollment sizes in the baseline year (not subject to WWC criteria, but on which we did nevertheless obtain strong baseline equivalence). The extent to which the Lone Star schools and the selected comparison schools were similar in terms of demographic characteristics—which, other than baseline-year enrollment size, we did not consider in the matching process—is discussed later in this chapter.

Exhibit 6.1. Baseline Statistics for the Year 2, Subsample 1 Analysis Sample

Measure	Lone Star schools (*n* = 13)	Comparison schools (*n* = 91)	*p* value	Hedges' *g*
Percentage of 2020 AP CSP exam takers earning a 3, 4, or 5 on the exam	65.30% (21.08)	65.65% (19.81)	0.953	-0.017
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2020 AP CSP exam	1.19% (0.90)	1.23% (0.97)	0.897	-0.038
Percentage of students in grades 9–12 earning a 1 or 2 on the 2020 AP CSP exam	0.79% (1.38)	0.73% (1.21)	0.885	0.043
Percentage of students in grades 9–12 taking the 2020 AP CSP exam	1.98% (1.85)	1.96% (1.83)	0.977	0.008

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from independent-samples *t* tests; they are provided for parallelism with the outcome analysis results. Hedges' *g* in the last column is calculated as explained in previous chapters.

Subsample 2 Matching and Baseline Equivalence

As in the Year 1 analysis, Subsample 2—composed of schools that did not have AP CSP exam takers in 2020 but had takers of *other* subjects' AP exams in 2020, with unsuppressed data—had two blocks of schools, differing based on the profiles of the Lone Star schools in the *outcome year*—in this case 2022.

Block 2a was based on 16 Subsample 2 Lone Star schools that had AP CSP exam takers, with unsuppressed data, in 2022; they were matched with comparison schools of the same profile, of which there were 157 across the five states. Block 2b was based on five Subsample 2 Lone Star schools that did not have AP CSP exam takers in 2022 (in addition to not having had AP CSP exam takers in 2020). They were matched with comparison schools that did not have *suppressed* AP CSP data in 2022; that is, the potential comparison schools either had *no* AP CSP exam takers in 2022 (like the Block 2b Lone Star schools) *or, unlike* the Block 2b Lone Star schools, had unsuppressed 2022 AP CSP exam data.[5] Because unselected schools in the Block 2a pool met the criteria to be considered as matches in Block 2b, we conducted the matching for the two blocks sequentially, with Block 2a first. Potential comparison schools that were not selected for Block 2a were added to the matching pool for Block 2b, which then consisted of 1,489 schools. For both Subsample 2 blocks, we matched on their average 2020 AP exam performance across all AP subject areas[6] as proxy measures corresponding to the AP CSP outcomes (see Chapter 5), along with enrollment size in grades 9 through 12.

For Block 2a, we experimented with selecting up to seven comparison schools as the number of comparison schools to be matched with each Lone Star school. We decided on four as the optimal number (balancing strong baseline equivalence results with statistical power), resulting in a total of 64 comparison schools, 47 of which were in Texas. For Block 2b, we experimented with selecting up to 10 comparison schools per Lone Star school and settled on six as a ratio offering good baseline equivalence and statistical power, resulting in a total of 30 comparison schools, 17 of which were in Texas.[7] Exhibit 6.2 shows the states of the selected comparison schools in each block.

Exhibit 6.2. States of the Year 2, Subsample 2 Comparison Schools

Block	Texas	Georgia	Massachusetts	New York	Wisconsin	Total
2a	47	6	6	3	2	**64**
2b	17	0	4	5	4	**30**
Subsample 2 Total	**64**	**6**	**10**	**8**	**6**	**94**

The within-block baseline equivalence statistics for each of the Subsample 2 blocks, which were used to select the comparison schools within each block, are provided in Appendix Exhibit E-2.

[5] See footnote 15 in Chapter 5 for an explanation for why we did not limit the comparison schools in this block to those that, like the Lone Star schools in the block, had no AP CSP exam takers in 2022. In the matching process, we purposely remained blind to whether or not the potential comparison schools had AP CSP exam takers in 2022.
[6] See footnote 16 in Chapter 5 for additional detail about why we included all AP subject areas.
[7] As noted in an earlier footnote, we purposely remained blind, during the matching, to whether the schools in the Block 2b comparison pool had AP CSP exam takers in 2022. It turned out that of the 30 selected schools, none had AP CSP exam takers in 2022.

The combined-block baseline equivalence statistics, calculated as described in Chapter 5, are presented in Exhibit 6.3. As shown in the last column, the Hedges' *g* for each baseline measure was less than (an absolute value of) 0.25 standard deviations, which meets WWC criteria provided that the baseline measure is included as a covariate in the impact analysis for its corresponding outcome measure (which it was).

Exhibit 6.3. Baseline Statistics for the Year 2, Subsample 2 Analysis Sample

Measure	Lone Star schools (*n* = 21)	Comparison schools (*n* = 94)	Regression coefficient (β_{Trt})	*p* value	Hedges' *g*
Average percentage (across all 38 AP subject area exams) of 2020 exam takers earning a 3, 4, or 5	33.38% (16.11)	36.00% (15.55)	-2.27%	0.549	-0.144
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 3, 4, or 5 in 2020	0.28% (0.26)	0.29% (0.29)	-0.02%	0.800	-0.061
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 1 or 2 in 2020	0.47% (0.36)	0.44% (0.32)	0.02%	0.815	0.056
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 taking a 2020 AP exam	0.74% (0.54)	0.73% (0.54)	0.00%	0.994	0.002

Note. In the first two columns, each cell contains the unadjusted mean and (in parentheses) standard deviation of the baseline measure. In the third column, the regression coefficient is the coefficient on the Lone Star indicator from the regression of the baseline measure on the Lone Star indicator and matching block indicator (i.e., the block-adjusted baseline mean difference). The Hedges' *g* value results from the inclusion of the regression coefficient as the numerator (i.e., the block-adjusted mean difference) in the Hedges' *g* formula.

In summary, for both of the two subsamples, we succeeded in selecting a well-matched set of comparison schools—equivalent to the Lone Star schools at baseline on the outcomes to be examined—for the 2022 analysis of program impacts on AP CSP exam performance and participation.

Statistical Estimation Strategy

The statistical estimation strategy we used to estimate the effect of the Lone Star program on the 2022 AP CSP exam performance and participation outcomes (i.e., impact after two years of program implementation) was the same as that used for the Year 1 analysis (see Chapter 5). Specifically, we began by running a separate ordinary least squares (OLS) regression for each of the four outcomes in each of the two subsamples, with a treatment indicator as the key predictor of interest, a baseline measure of the outcome as a key covariate to adjust for minor

baseline differences, a block indicator if needed, and school demographic characteristics as additional covariates.[8] The school demographic covariates included in the regression models were the same as those in the Year 1 analysis except that they (like the outcome measures) were from the 2021–22 school year;[9] these data were drawn from the Common Core of Data,[10] except for the economically disadvantaged variable for Massachusetts and Georgia schools, which came from state data.[11]

Appendix Exhibits E-3 and E-4 provide descriptive statistics on all of the covariate measures for the Subsample 1 Lone Star and comparison schools, and Appendix Exhibits E-5 and E-6 provide the covariate measure descriptives for the Lone Star and comparison schools in Subsample 2. As was the case in the Year 1 sample of schools (see Chapter 5), there were some notable demographic differences between the Lone Star schools and the selected comparison schools, as indicated by the effect sizes shown in the rightmost column of each table. These differences were along the same lines as those observed in Year 1:

- In both Subsample 1 and Subsample 2, Lone Star schools were much more likely than the comparison schools to be charter schools.
- In Subsample 1, Lone Star schools had much higher percentages of Hispanic students than comparison schools, on average, and somewhat lower percentages of female students. They were also more likely to be located in cities and less likely to be located in suburbs.
- In Subsample 2, Lone Star schools had notably lower percentages of female students, on average, and slightly higher percentages of Hispanic students. They were also notably more likely to be in cities, and much less likely to be in rural areas or towns.

[8] See Chapter 5 for detail about why we conducted a separate regression for each subsample and how the models for the two subsamples differed from one another.
[9] See footnote 19 in Chapter 5 for an explanation for why we used the outcome year (in this chapter, 2021–22) rather than the 2019–20 baseline year as the year for all school demographic characteristics.
[10] Common Core of Data: https://nces.ed.gov/ccd/elsi/tableGenerator.aspx.
[11] For schools in all states except Massachusetts and Georgia, the economically disadvantaged covariate was the school percentage of students in the free and reduced-price lunch (FRL) program based on counts provided in the Common Core of Data (CCD). However, FRL data were not available in CCD for Massachusetts schools in any of our analysis years, and they were not available for Georgia schools in the 2021–22 year because they did not meet quality standards. For Massachusetts schools, we instead used percentages of "low income" students, based on counts in data files at https://profiles.doe.mass.edu/statereport/selectedpopulations.aspx. (Massachusetts apparently switched from "economically disadvantaged" in 2020–21 to "low income" in 2021–22.) For Georgia schools, we used school 2021–22 FRL percentages from a data file downloaded from https://gosa.georgia.gov/dashboards-data-report-card/downloadable-data, except for two schools whose 2021–22 FRL values seemed implausible, based on a comparison with their FRL values in other study years; particularly in light of the CCD data quality concerns about Georgia's 2021–22 FRL, we decided to treat FRL as missing for these two schools. (Both schools were in the same school district, and they were the only two schools in that district in our comparison sample.) Because of the difference between states on this variable, we standardized the values within state. So as not to exclude the two schools in Georgia that were missing economically disadvantaged data, along with one school in New York also missing economically disadvantaged data, we imputed the missing values as zero (i.e., the within-state standardized mean) and included an indicator of missingness in the model. No other variables had missing data.

As previously mentioned, our statistical models controlled for these and other school characteristics.

After conducting the regressions, we calculated the effect size of the estimated intervention effect for each of the four measures in each of the two subsamples, using the *adjusted* mean difference in the *outcome* year, namely the β_{Trt} coefficient from the regression, as the numerator in the Hedges' *g* formula. We then conducted a meta-analysis combining the estimates from the two subsamples to yield a single overall estimate for each outcome. (See Chapter 5 for further detail on how the meta-analysis was conducted.)

Results

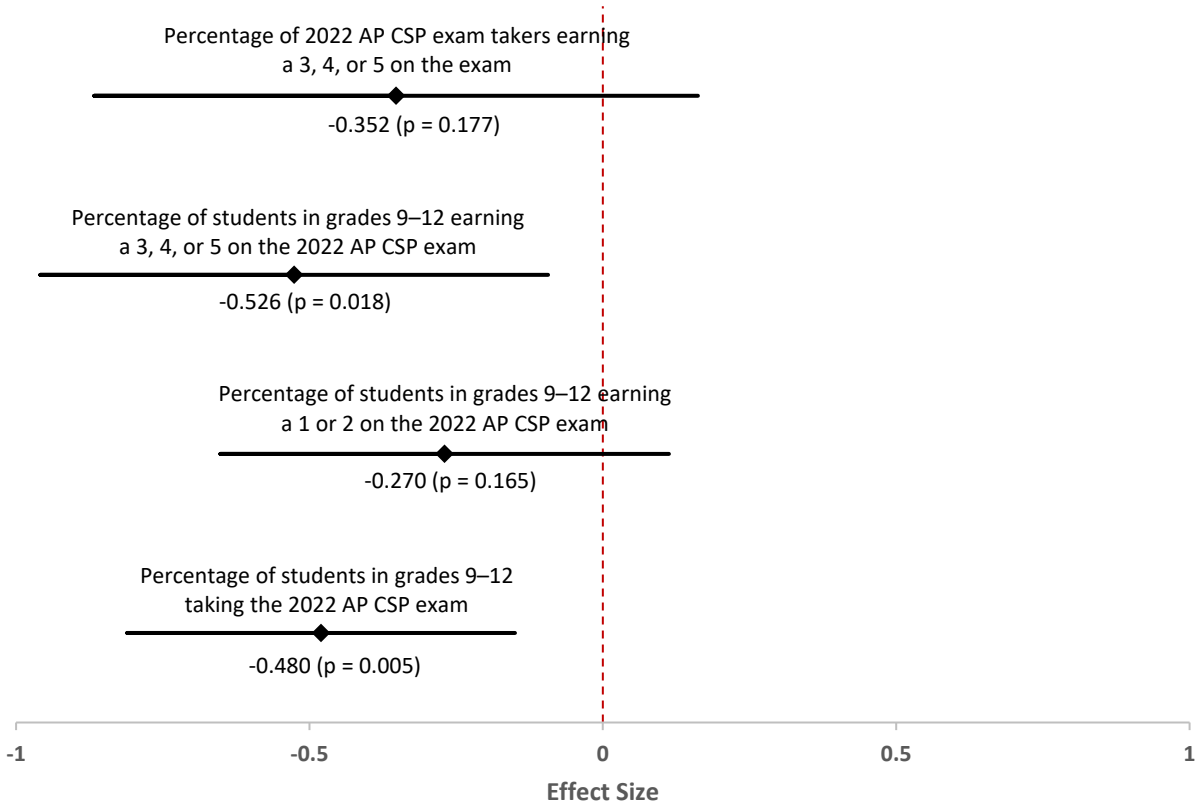
In this section we first present the results separately for each of the two subsamples and then present the meta-analysis (combined) results.

Subsample 1 and 2 (OLS-Based) Results

Exhibits 6.4 and 6.5 show the estimated effect sizes for the second-year effects of the Lone Star AP CSP program on the four outcome measures for Subsample 1 and Subsample 2, respectively. Exhibit 6.6 provides additional details, including the unadjusted (i.e., descriptive) means and standard deviations of each outcome measure, the regression-adjusted mean difference (i.e., the coefficient on the Lone Star indicator from the OLS regression), the *p* value of the coefficient from the regression, and the calculated Hedges' *g* effect size (identical to the point estimates shown in Exhibits 6.4 and 6.5). The full regression results are provided in the appendix (Exhibits E-7 and E-8 for each subsample, respectively).

For schools in Subsample 1 (i.e., those that had AP CSP exam takers in the baseline year), the Lone Star AP CSP program appears to have had a statistically significant effect on two of the Year 2 outcomes: percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2022 AP CSP exam and percentage of students in grades 9–12 taking the exam (Exhibit 6.4). Both effects are negative and moderate in size—about one half of a standard deviation. Most likely, the two effects are related, given that both have the grades 9–12 enrollment as the denominator for the percentage: If fewer students in the school are taking the exam, it is likely that fewer students in the school will earn a qualifying score. A logical corollary would be that fewer students in the school would earn a nonqualifying score as well, and that is what the results show for the third outcome—percentage of students in grades 9–12 earning a 1 or 2 on the 2022 exam—although that effect is not statistically significant. The estimate for the other outcome—percentage of 2022 AP CSP exam takers earning a 3, 4, or 5 on the exam—also is negative but not statistically significant.

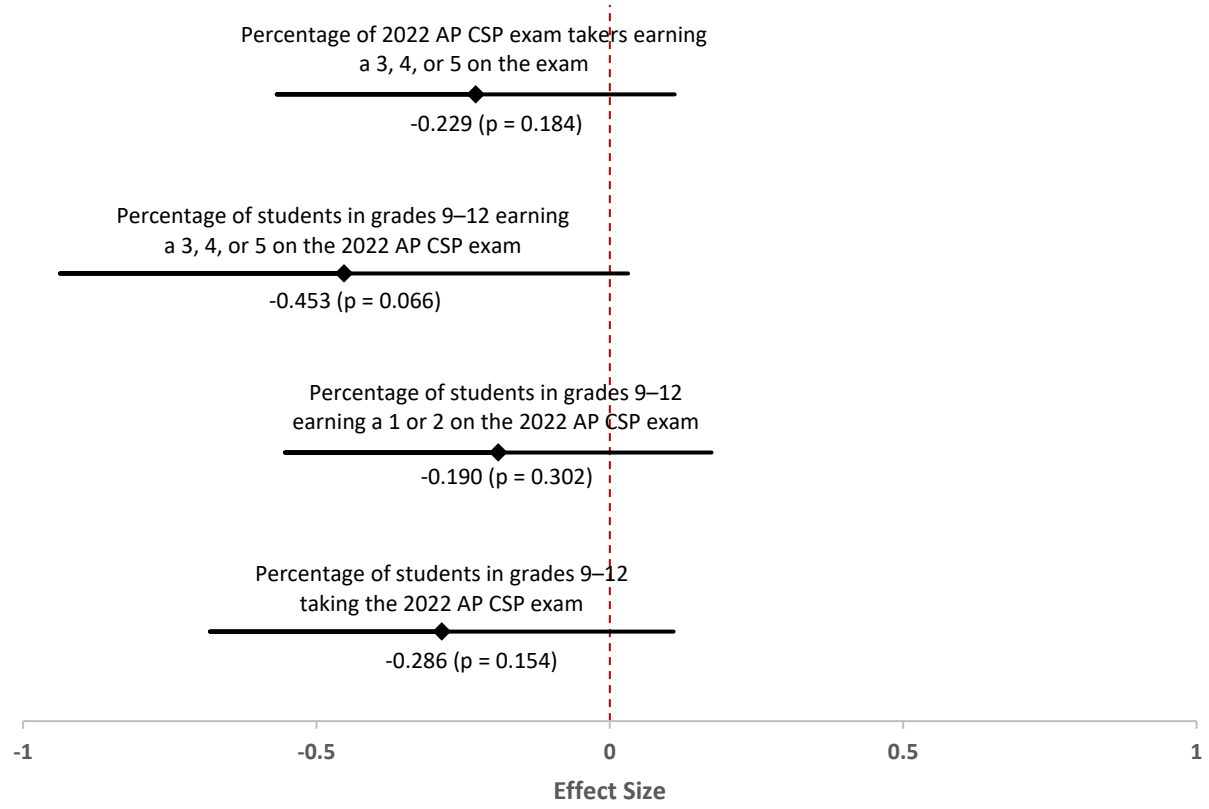
Exhibit 6.4. Effect Size Chart for the Year 2, Subsample 1 Analysis



Note. *N* = 104 (13 Lone Star schools and 91 comparison schools). Effect size point estimates are represented by the diamonds; the horizontal lines show 95% confidence intervals. Confidence intervals that do not intersect with an effect size of 0 (represented by the dotted red line) indicate a statistically significant ($p < .05$) effect.

For Subsample 2 schools, which did not have AP CSP exam takers in the baseline year, there were no statistically significant effects for any of the four outcomes (Exhibit 6.5). However, as in Subsample 1, the estimates for all four outcomes are negative. The effect for the percentage of students in grades 9–12 earning a 3, 4, or 5 is moderate in size, at nearly half of a standard deviation, and only barely misses statistical significance at the $p < .05$ level.

Exhibit 6.5. Effect Size Chart for the Year 2, Subsample 2 Analysis



Note. $N = 115$ (21 Lone Star schools and 94 comparison schools). Effect size point estimates are represented by the diamonds; the horizontal lines show 95% confidence intervals. Confidence intervals that do not intersect with an effect size of 0 (represented by the dotted red line) indicate a statistically significant ($p < .05$) effect.

Exhibit 6.6. Year 2 Outcome Statistics for Both Subsamples

Measure	Lone Star schools (unadj.)	Comparison schools (unadj.)	Adj. mean difference (β_{Trt})	*p* value	Hedges' *g*
Subsample 1 (13 Lone Star schools and 91 comparison schools)					
Percentage of 2022 AP CSP exam takers earning a 3, 4, or 5 on the exam	42.56% (18.00)	62.23% (26.52)	-9.11% (6.70)	0.177	-0.352
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2022 AP CSP exam	1.12% (1.42)	1.29% (1.09)	-0.60% (0.25)	0.018	-0.526
Percentage of students in grades 9–12 earning a 1 or 2 on the 2022 AP CSP exam	1.41% (1.15)	1.07% (2.09)	-0.54% (0.39)	0.165	-0.270
Percentage of students in grades 9–12 taking the 2022 AP CSP exam	2.53% (2.10)	2.36% (2.61)	-1.24% (0.43)	0.005	-0.480
Subsample 2 (21 Lone Star schools and 94 comparison schools)					
Percentage of 2022 AP CSP exam takers earning a 3, 4, or 5 on the exam	18.44% (23.23)	25.89% (28.72)	-6.40% (4.79)	0.184	-0.229
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2022 AP CSP exam	0.46% (0.80)	0.73% (1.58)	-0.67% (0.36)	0.066	-0.453
Percentage of students in grades 9–12 earning a 1 or 2 on the 2022 AP CSP exam	2.12% (5.61)	1.21% (2.62)	-0.64% (0.62)	0.302	-0.190
Percentage of students in grades 9–12 taking the 2022 AP CSP exam	2.58% (5.81)	1.94% (3.84)	-1.23% (0.85)	0.154	-0.286

Note. In the first two columns, each cell contains the unadjusted mean and (in parentheses) standard deviation of the outcome measure. The third column contains the regression coefficient on the Lone Star indicator from the impact analysis model (i.e., the adjusted mean difference) and (in parentheses) the standard error of the coefficient; the *p* value associated with the coefficient is in the fourth column. The Hedges' *g* value in the final column results from the inclusion of the regression coefficient as the numerator (i.e., the mean difference) in the Hedges' *g* formula.

The negative results of the statistical analysis may seem surprising and difficult to comprehend when compared to the descriptive results (first two columns of Exhibit 6.6), particularly for the percentage of students taking the AP CSP exam. For Subsample 1, the descriptive results show the Lone Star schools as having a higher percentage of students taking the exam in 2022

(2.53%) than the comparison schools (2.36%), even though they started off similarly in 2020, when 1.98% of students at Lone Star schools and 1.96% of students at comparison schools took the exam (Exhibit 6.1, earlier). In other words, the percentage of students taking the exam increased by 0.55 points at Lone Star schools while the comparison school percentage increased only by 0.40 points. And in Subsample 2, the descriptive results show that the Lone Star schools had 0.64% more students than comparison schools taking the AP CSP exam in 2022 (2.58% at Lone Star schools vs. 1.94% at comparison schools).

To what can we attribute the difference between the descriptive results and the statistical results? The answer lies mainly in the inclusion of school demographic variables as covariates in the statistical analyses, in combination with the fact that the Lone Star schools and the comparison schools were dissimilar on several of these variables. The demographic variables—or at least some of them—are strongly related to schools' 2022 AP CSP exam participation rates. In Subsample 1, the demographic covariates explain an additional 13.5% of the variance in the participation outcome compared to a model that includes only the treatment and baseline indicators; in Subsample 2, they explain an additional 20% of the variance compared to a model without them. And in both subsamples, the inclusion of the demographic covariates turns the treatment effect from positive to negative. (See Appendix Exhibits E-9 and E-10.)

The most influential variable, among the demographic characteristics, appears to be the percentage of school enrollment that is female. In both subsamples, the proportion of total enrollment that is female is a highly significant ($p < .001$) *negative* predictor of AP CSP exam taking. All else being equal, for each percentage point larger a school's female population is (e.g., 49% vs. 48% female), the percentage of grades 9–12 students taking the 2022 AP CSP exam drops by 0.14 percentage points in Subsample 1 and by 0.25 percentage points in Subsample 2. These percentage point declines may sound small, but given that the percentages of students taking the exam are also small—in the 2% total range—they are actually quite substantial. Because Lone Star schools, on average, had substantially lower percentages of female students than comparison schools, the statistical expectation based on their student composition would be that the Lone Star schools would have *even larger* increases in AP CSP exam participation than they actually did, such that, relative to the comparison schools, the net effect (i.e., the *adjusted* mean difference) is negative.[12]

[12] These findings may lead to speculation as to whether the results would have been different if we had included percent female as a matching variable and thereby obtained a sample of comparison schools that, on average, were similar to the Lone Star schools on percent female as well as on the outcomes as baseline (assuming this would have been possible, which is not a given). The inference based on the results from the actual sample is that the Lone Star effect in the statistical model using a more gender-balanced sample would still have been significantly negative, but that the *descriptive* results would have looked less favorable to the Lone Star schools.

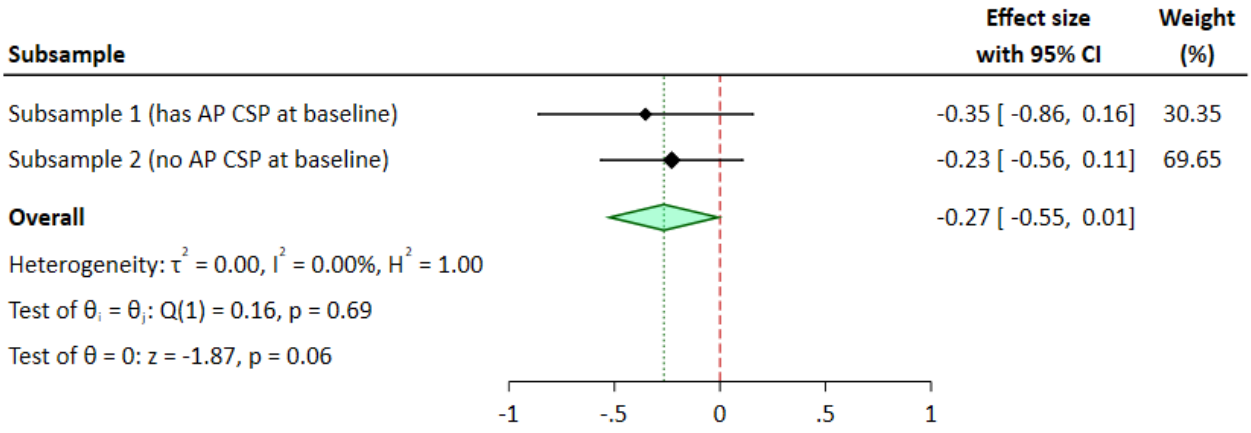
Meta-Analysis Results

Exhibits 6.7 through 6.10 display the Year 2 meta-analysis results for each of the four outcome measures, respectively. As the exhibits show, there was a statistically significant, moderately sized negative overall effect for two of the 2022 outcomes: percentage of students in grades 9–12 earning a qualifying score on the AP CSP exam (Exhibit 6.8) and percentage of students in grades 9–12 taking the AP CSP exam (Exhibit 6.10). The overall estimates for the other two outcomes—percentage of AP CSP exam takers earning a qualifying score (Exhibit 6.7) and percentage of students in grades 9–12 earning a nonqualifying score on the AP CSP exam (Exhibit 6.9)—were also both negative but were only borderline significant ($p < .10$).

As discussed in the presentation of the Subsample 1 results, the negative findings for the two performance outcomes that have grades 9–12 enrollment as the denominator for the percentage calculations are likely explained by the negative finding for the participation outcome: If fewer students in a school take the AP CSP exam, the likelihood, all else being equal, is that fewer students in the school earn either a qualifying score or a nonqualifying score. However, the suggestive negative estimate for the first performance outcome—percentage of AP CSP exam takers earning a qualifying score—cannot be explained by the participation finding.

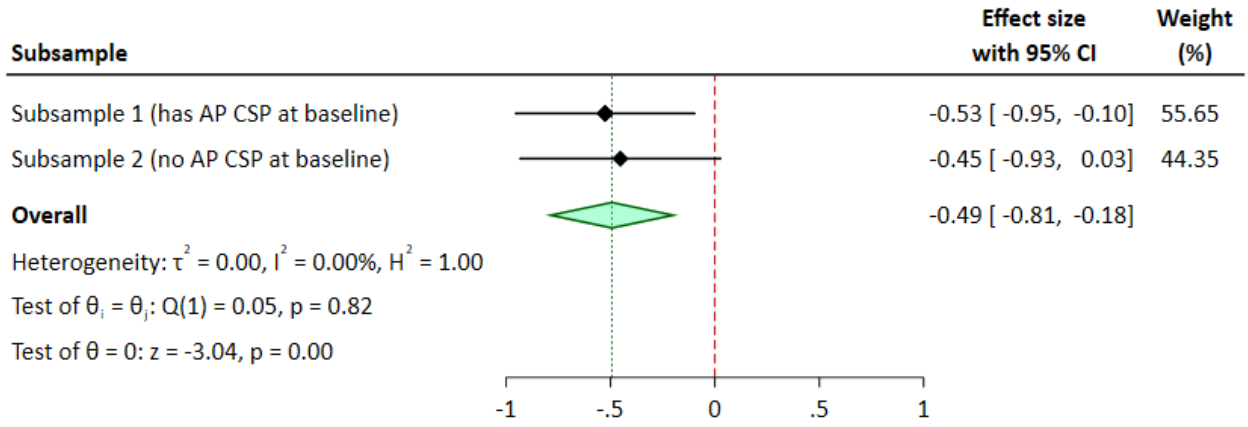
The meta-analysis results also indicate that there was no significant variability across the two subsamples for any of the four outcomes, and in no case was the effect for either subsample significantly different than the overall effect. In fact, the results for the two subsamples are remarkably similar to one another.

Exhibit 6.7. Meta-Analysis Plot for Percentage of 2022 AP CSP Exam Takers Earning a 3, 4, or 5 on the Exam



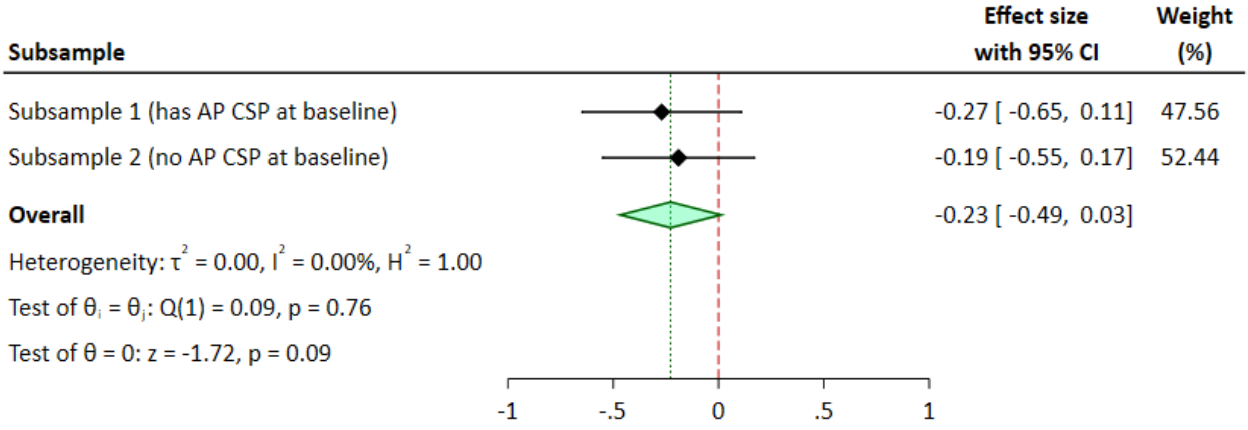
Note. See box in Chapter 5 titled "How to Read the Meta-Analysis Figures."

Exhibit 6.8. Meta-Analysis Plot for Percentage of Students in Grades 9–12 Earning a 3, 4, or 5 on the 2022 AP CSP Exam



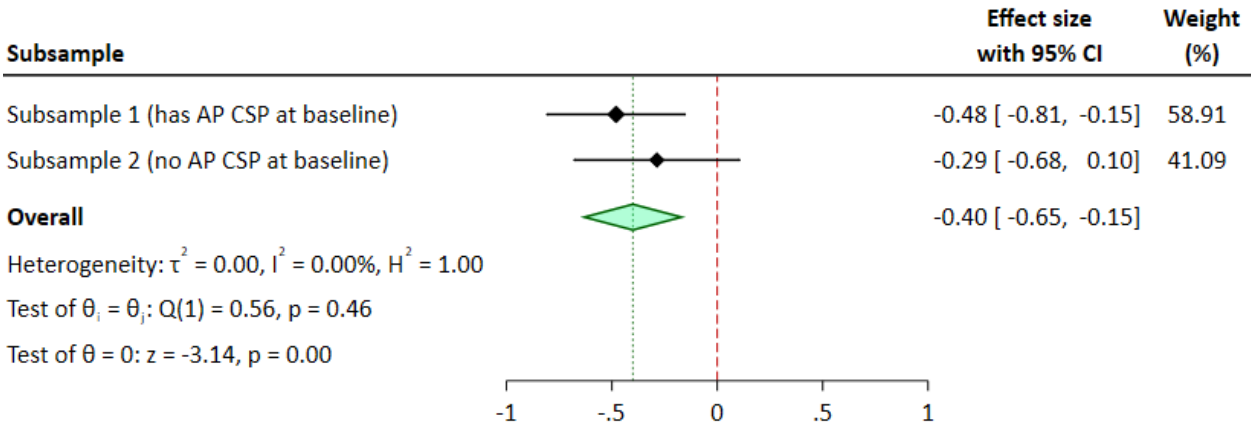
Note. See box in Chapter 5 titled "How to Read the Meta-Analysis Figures." For test of θ, $p = 0.0024$.

Exhibit 6.9. Meta-Analysis Plot for Percentage of Students in Grades 9–12 Earning a 1 or 2 on the 2022 AP CSP Exam



Note. See box in Chapter 5 titled "How to Read the Meta-Analysis Figures."

Exhibit 6.10. Meta-Analysis Plot for Percentage of Students in Grades 9–12 Taking the 2022 AP CSP Exam



Note. See box in Chapter 5 titled "How to Read the Meta-Analysis Figures." For test of θ, $p = 0.0017$.

WHAT ABOUT THOSE OMITTED SCHOOLS?

As discussed earlier in this chapter, 34 of the 37 Lone Star AP CSP schools were included in the Year 2 analyses. The other three schools could not be included because they lacked usable baseline measures: one school had no AP exam takers at all in 2020 and the other two schools had fewer than five AP CSP exam takers in 2020 and thus their AP CSP exam score data were suppressed.

All three of these omitted schools, however, had more than five AP CSP exam takers in 2022, the outcome year studied in this chapter. This raises the question of whether the exclusion of these schools might have biased the results, particularly given the finding of an apparent negative effect of the Lone Star AP CSP program on AP CSP exam participation. If these schools, which all increased the number of AP CSP exam takers from 2020 to 2022, could be included in the analyses, might the participation finding or any of the others have been different?

To consider this, we conducted exploratory analyses that added in two of the three schools. There was no simple way to include the school that had no AP exam takers at all in 2020, but for the two that had suppressed AP CSP exam data, we could include them as an additional block in Subsample 2 by using their all-AP-subjects-combined percentages as their baseline measures, as we did for the schools in Blocks 2a and 2b. Out of a pool of 91 potential comparison schools that had the same profile as the two Lone Star schools—suppressed AP CSP exam scores in 2020, unsuppressed all-AP-subjects-combined scores in 2020, and unsuppressed AP CSP exam scores in 2022—we identified a set of 14 comparison schools (including 11 in Texas) that were equivalent to the two Lone Star schools on the four baseline measures (i.e., baseline mean differences of less than an absolute value of 0.25 standard deviations).[a] We then reran the Subsample 2 analyses with these additional 16 schools included (along with an additional indicator for the third block) and, following that, we conducted a new meta-analysis for each outcome to combine the existing Subsample 1 results with the results of the expanded, exploratory Subsample 2 analysis.

The exploratory Subsample 2 results were essentially the same as the main Subsample 2 results; all remained notably negative. For the three outcomes with grades 9–12 as the denominator for the percentage, the effects in the exploratory analysis were a tiny bit smaller (closer to zero) than in the main analysis, but none changed significance levels; the effect for the percentage of exam takers earning a qualifying score became slightly larger (more negative) but also did not change significance levels. In the meta-analysis, on the other hand, two of the results did change significance levels. The negative effect for the percentage of AP CSP exam takers earning a qualifying score, which was only borderline significant in the main analysis, became slightly more negative (point estimate -0.28 SD) and significant ($p < .05$). The negative effect for the percentage of students in grades 9–12 earning a nonqualifying score, which also was borderline significant in the main analysis, became slightly less negative (point estimate -0.18 SD) and nonsignificant. The effects for the other two outcomes, including the participation outcome, remained negative, of moderate magnitude, and highly significant. Thus, the exploratory results including two additional Lone Star schools are no more favorable for the effectiveness of the program, including AP CSP exam participation rates, than the main results.

[a] Stata's *kmatch* failed to identify a set of matched comparison schools (using either Mahalanobis distance matching or propensity score matching), so we instead used the hand-matching tool (Holtzman, 2024) described in Chapter 7.

Summary and Discussion

In summary, we found that after two years of implementation, the Lone Star AP CSP program appears to have had a statistically significant negative effect on the percentages of students in grades 9–12 that took the AP CSP exam, which in turn also caused fewer students in grades 9–12 to earn either a qualifying score or a nonqualifying score on the exam, relative to matched comparison schools and controlling for baseline measures and school demographics. For the percentage of AP CSP *exam takers* who earned a qualifying score, the overall effect of the Lone Star program was negative but only borderline significant.

Although descriptively the Lone Star schools had greater increases in AP CSP exam participation rates than did the comparison schools, when demographic characteristics of the schools are accounted for—especially the percentage of students who are female—the net effect of the program on Year 2 participation rates was negative.

As noted in Chapter 2, the school supports component of the Lone Star AP CSP program, which according to the logic model was theorized to be a key driver of increased rates of participation in AP CSP, did not meet fidelity-of-implementation criteria in Year 2. Even so, it is still somewhat surprising to observe a *negative* effect of the program on AP CSP participation, as opposed to a merely null effect. Moreover, students' participation in AP CSP in Year 2 might be more reasonably expected to be linked with the school supports component in Year 1 rather than in Year 2—given that AP CSP is typically a year-long course that students would have enrolled in at the *start* of Year 2 (or perhaps even signed up for at the end of Year 1)—and the

component *did* meet fidelity criteria in Year 1. However, as noted in the Chapter 5 summary, the specific elements of the school supports component that met the fidelity criteria in Year 1 were those that examined provision of resources to the schools, rather than uptake.

In addition, while the school supports component might have been the theorized key driver of participation in the AP CSP *course*, the participation outcome we actually examine is participation in the AP CSP *exam*. Participation in the exam might be more related to the *student* supports component (e.g., provision of AP exam fee subsidies[13]), which according to the evaluation's criteria was implemented with fidelity in Year 2, but it could also be related to the teacher supports component, which was not.

Factors external to the program could help explain the findings. The second year of implementation was the 2021–22 school year, when the negative effects of the COVID-19 pandemic on students nationwide were becoming more apparent. It is possible that in response, schools de-emphasized enrollment in elective classes such as AP CSP in favor of more instruction in core subject areas or more time devoted during the school day to students' social-emotional well-being. It is not obvious why this would have been more the case in the Lone Star AP CSP schools than in the matched comparison schools, but it could have been something about the Texas context, given that only some of the comparison schools were located in Texas. In Subsample 1, only about one third of the comparison schools were in Texas, whereas in Subsample 2, more than two thirds of the schools were in Texas; in combination with the fact that the negative effect on participation was more prominent in Subsample 1 than in Subsample 2, this lends some support to the Texas theory. Further exploration of the theory, in the form either of researching Texas's post-pandemic educational priorities and responses relative to those of the other states in which comparison schools were located, or of additional empirical analysis with the currently available data (e.g., reanalysis with comparison schools in Texas only), could be worthwhile. (See the addendum to Chapters 5–7 that follows Chapter 7 for some exploration of the latter.)

The borderline-significant negative finding for the program's apparent effect on the Year 2 percentage of exam takers earning a qualifying score on the exam, meanwhile, cannot be explained by the finding of reduced participation. Notably, it is at odds with the large positive significant effects of the program on AP CSP teachers' Year 2 instructional practice reported in Chapter 4. Although we can only speculate on reasons for the discrepancy, among the possibilities are that (a) teachers misrepresented their instructional practice on the survey, whether intentionally or not; (2) the sample of Lone Star AP CSP teachers included in the Year 2 instructional practices analysis was at best a small subset of the teachers in the Lone Star AP

[13] As discussed in Chapter 2, the AP exam fee subsidies were provided only to schools that requested them, and hardly any schools did (Exhibit 2.3). We do not know why more of the schools did not request them.

CSP schools included in the Year 2 AP performance analysis; or (3) the instructional practices that showed increased use are not actually positively associated with improved performance on the AP CSP exam by students.

Chapter 7. Impact on Student AP CSP Performance and Participation After Three Years

This chapter examines whether the program achieved its aims for AP CSP performance and participation (see Exhibit 2.1 in Chapter 2) during its third year of operation, the 2022–23 school year. The research question addressed is as follows:

6. What was the impact of three years of participation in the Lone Star AP CSP program, compared to business-as-usual AP CSP participation, on high schools' rates of (a) AP CSP exam takers receiving a qualifying score on the AP CSP exam, and (b) students taking the AP CSP exam?

Chapters 5 and 6 looked at the impact after one and two years of implementation, respectively. This is the final chapter on the impact of the program on student AP CSP performance, given that the program as evaluated by AIR lasted for three years.

As described in the preceding chapters, the student impact study used a quasi-experimental matched comparison group design and an analysis-of-covariance (regression) estimation model to examine the effect of Lone Star AP CSP on the performance and participation outcomes, followed by meta-analysis to combine the estimates from two samples of schools (those that had AP CSP exam takers in the baseline year and those that did not).

KEY FINDINGS

- After three years of implementation, the Lone Star AP CSP program did not have a significant effect on schools' percentages of AP CSP exam takers that received a qualifying score on the exam.
- The program also did not have a significant effect after three years on schools' percentages of students in grades 9–12 that received a qualifying score on the AP CSP exam, but it did have a significant effect on schools' percentages of students in grades 9–12 that received a *non*qualifying score. This effect was negative, which is in the desired direction.
- The program also had a significant negative effect on schools' percentages of students in grades 9–12 that took the AP CSP exam in the third year of implementation.

The remainder of this chapter is organized in the same way as Chapter 6: a brief recap of the outcome measures and data sources, information on the schools in the analysis, a brief recap of the estimation strategy, detailed presentation of results, and finally a brief summary and discussion.

Outcome Measures

The outcome measures examined in the analysis of the impacts of the third year of Lone Star AP CSP—the 2022–23 school year—are the same as those described in Chapters 5 and 6, except that they are based on the AP CSP exam administered in spring 2023:[1]

- the percentage of AP CSP exam takers earning a 3, 4, or 5 on the exam *(key measure of student performance)*;
- the percentage of students in grades 9–12 earning a 3, 4, or 5 on the AP CSP exam *(supplementary measure #1 of student performance)*;
- the percentage of students in grades 9–12 earning a 1 or 2 on the AP CSP exam *(supplementary measure #2 of student performance)*; and
- the percentage of students in grades 9–12 taking the AP CSP exam *(measure of student participation)*.[2]

The data sources for the outcome measures and their corresponding baseline measures were also the same as in the first- and second-year impact analyses:

- The AP data for schools in Texas (including the Lone Star schools) came from datasets prepared by the College Board.
- The AP data for schools in four other states—Georgia, Massachusetts, Wisconsin, and New York—came from datasets posted publicly on those states' websites.[3]
- Enrollment data, used to calculate the percentages for the participation measure and the two supplementary performance measures, came from the websites of the five states.

See Chapter 5 for further detail.

Lone Star School Sample and Selection of Comparison Schools

As noted in earlier chapters, 37 Texas schools were recruited by NMSI to participate in the Lone Star AP CSP intervention, and 36 of them were eligible for inclusion in the AP exam impact analyses (in all outcome years). Of these 36 Lone Star schools, 32 were included in the analyses of the impact of the program in the 2022–23 school year on the four outcomes of interest. As in the Year 2 analyses, two schools could not be included in the Year 3 analysis because their 2020 (baseline year) AP CSP scores were suppressed due to small cell sizes (i.e., fewer than five exam

[1] For the three measures using grades 9–12 enrollment as the denominator for the percentage calculations, school enrollments were from the same school year (2022–23).
[2] See Chapter 5 for further detail about the four measures, including discussion of how they are mathematically related to one another and why this is potentially important in the interpretation of results.
[3] See footnote 6 in Chapter 5 for further explanation of and rationale for why we allowed some comparison schools to be located outside of Texas.

takers). One school could not be included because it had fewer than five AP CSP exam takers in 2023 and thus its 2023 score data were suppressed. Finally, one school was omitted because, despite having had at least five AP CSP exam takers in the baseline year (2020), it had *no* AP CSP exam takers in 2023. In theory this school could have been included in analyses with its 2023 outcome percentages coded as 0s, but because it was the only Lone Star school with this profile, there would have been statistical complications in identifying matched comparison schools and demonstrating baseline equivalence.[4]

With 32 schools, the Year 3 analyses included more Lone Star schools than were included in the Year 1 analyses (28), but fewer than were included in the Year 2 analyses (34). See the box for additional information on the Lone Star schools' different patterns of AP CSP exam taking across the four years covered by the evaluation (i.e., the baseline year plus the three outcome years).

[4] As described in earlier chapters, demonstration of baseline equivalence for continuous measures is done using Hedges' *g*, which is a standardized mean difference. The standardization relies on both intervention and comparison schools having a standard deviation around the mean, which requires multiple schools in each group. That is, an *n* of 1 does not have a standard deviation.

LONE STAR SCHOOLS' CROSS-YEAR PATTERNS OF AP CSP EXAM TAKERS: BASELINE YEAR (BL), YEAR 1 (Y1), YEAR 2 (Y2), AND YEAR 3 (Y3)

Cross-year pattern (BL, Y1, Y2, Y3)	Number (percent) of schools	Subsample/block			Which years included in AP analysis
		Y1	Y2	Y3	
●●●●	12 (32%)	1	1	1	All
●●●○	1 (3%)	1	1	1	Y1, Y2[a]
○●●●	8 (22%)	2a	2a	2a	All
○●●○	2 (5%)	2a	2a	2b	All
○●○○	1 (3%)	2a	2b	2b	All
○○●●	2 (5%)	2b	2a	2a	All
○○○●	2 (5%)	2b	2b	2a	All
○⊗●●	2 (5%)	-	2a	2a	Y2, Y3
○⊗●○	1 (3%)	-	2a	2b	Y2, Y3
○⊗●⊗	1 (3%)	-	2a	-	Y2
○⊗○○	2 (5%)	-	2b	2b	Y2, Y3
⊗●●●	1 (3%)	-	-	-	None
⊗⊗●●	1 (3%)	-	-	-	None
●●●	1[b] (3%)	-	-	-	None
TOTAL	**37 (100%)**				

○ No AP CSP exam takers ⊗ 1 to 5 AP CSP exam takers (therefore scores suppressed)

● Five or more AP CSP exam takers (therefore scores unsuppressed)

[a] In theory this school could also have been in the Y3 analysis; see the main text of this chapter for an explanation of why it was not.

[b] This school could not be included in any year's analysis because it had no AP exam takers at all (in any subject) in the baseline year.

The table shows that of the 37 schools in the Lone Star AP CSP program:

- All 37 had AP CSP exam takers in at least one of the three years.
- 35 had at least five AP CSP exam takers in at least one of the three years.
- 22 had at least five AP CSP exam takers in all three years.
- 8 had at least five AP CSP exam takers in two of the three years.
- 5 had at least five AP CSP exam takers in one of the three years.

Subsamples and Matching Process

As in the analyses of impact after one and two years, the identification of matched comparison schools, as well as much of the subsequent analysis, occurred separately within each of two subsamples of schools:

- Subsample 1 consists of schools that had AP CSP exam takers in the baseline year (2020).
- Subsample 2 consists of schools that had did not have AP CSP exam takers in the baseline year (2020) but did have exam takers in *other* AP subjects that year. Subsample 2 had two blocks of schools, 2a and 2b, each with their own matching process, as explained further below.

For two of our three matching groups—Subsample 1 and Block 2a of Subsample 2—we again used the Stata module *kmatch* (Jann, 2017) to match Lone Star schools with comparison schools that were equivalent at baseline on key measures. (See Chapter 5 for further detail.) For Block 2b, however, we had to use a different matching method in Year 3. The reason for this and the alternate matching method we used are discussed below where we provide additional detail about this block.

Subsample 1 Matching and Baseline Equivalence

Subsample 1's single matching block ("Block 1") was built around 12 Lone Star schools that had AP CSP exam takers *and* unsuppressed AP CSP exam data in both 2020 and 2023.[5] They were matched with comparison schools of the same profile; there were 453 such schools in the five states for which we had data. Within this pool of potential comparison schools, we matched on the same four measures as in the Year 1 analysis (see Chapter 5) and, as in Year 1, we considered it essential to identify a single set of comparison schools to be used in the analyses for all four outcomes.

We experimented with selecting 1 to 15 comparison schools as the number of comparison schools to be matched with each Subsample 1 Lone Star school and decided on a 1:8 intervention-to-comparison ratio to provide both strong baseline equivalence on each of the four matching variables and statistical power for the analysis. Thus, from the pool of 453 potential comparison schools, we selected 96 for the analysis.

Of the 96 selected comparison schools, 37 were in Texas, 9 were in Georgia, 18 were in Massachusetts, 21 were in New York, and 11 were in Wisconsin. Exhibit 7.1 displays the

[5] These 12 schools were a subset of the 13 schools that were in Subsample 1 of the Year 1 and Year 2 analyses (Chapters 5 and 6, respectively). As in Year 2, we decided to rematch—that is, identify a new set of comparison schools—because some comparison schools selected as matches in the analysis of the one-year impact may not have had AP CSP exam takers in the third year (i.e., the 2022–23 school year), and because of the loss of one school.

baseline means, standard deviations, and standardized mean differences (i.e., Hedges' *g*) for the Subsample 1 Lone Star and selected comparison schools. The final column shows that for each of the four outcomes, the SMD on the baseline measure of the outcome was less than (an absolute value of) 0.05 standard deviations. As has been previously noted, baseline differences of less than 0.05 standard deviations meet WWC criteria regardless of whether the baseline measure is included as a covariate in the impact analysis, but we nevertheless did include the baseline measure as a covariate in the estimation model for its corresponding outcome (see "Statistical Estimation Strategy" section below). Exhibit F-1 in the appendix displays the equivalent information for schools' grades 9–12 enrollment sizes in the baseline year (not subject to WWC criteria, but on which we did nevertheless obtain strong baseline equivalence). The extent to which the Lone Star schools and the selected comparison schools were similar in terms of demographic characteristics—which, other than baseline-year enrollment size, we did not consider in the matching process—is discussed later in this chapter.

Exhibit 7.1. Baseline Statistics for the Year 3, Subsample 1 Analysis Sample

Measure	Lone Star schools (*n* = 12)	Comparison schools (*n* = 96)	*p* value	Hedges' *g*
Percentage of 2020 AP CSP exam takers earning a 3, 4, or 5 on the exam	68.66% (18.02)	68.77% (17.44)	0.983	-0.006
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2020 AP CSP exam	1.14% (0.92)	1.17% (0.92)	0.932	-0.026
Percentage of students in grades 9–12 earning a 1 or 2 on the 2020 AP CSP exam	0.41% (0.27)	0.41% (0.30)	0.989	-0.004
Percentage of students in grades 9–12 taking the 2020 AP CSP exam	1.55% (1.08)	1.58% (1.08)	0.939	-0.023

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from independent-samples *t* tests; they are provided for parallelism with the outcome analysis results. Hedges' *g* in the last column is calculated as explained in previous chapters.

Subsample 2 Matching and Baseline Equivalence

As in the prior two years' analyses, Subsample 2 had two separate blocks of schools because of differing Lone Star school profiles in the *outcome year*—in this case 2023. Block 2a was based on 14 Subsample 2 Lone Star schools that had AP CSP exam takers, with unsuppressed data, in 2023 (but not in 2020); they were matched with comparison schools of the same profile, of which there were 185 across the five states. Block 2b was based on six Subsample 2 Lone Star schools that did not have AP CSP exam takers in 2023 (in addition to not having had AP CSP exam takers in 2020). They were matched with comparison schools that did not have *suppressed* AP CSP data in 2023; that is, the potential comparison schools either had *no* AP CSP

exam takers in 2023 (like the Block 2b Lone Star schools) *or, unlike* the Block 2b Lone Star schools, had unsuppressed 2023 AP CSP exam data.[6] Because unselected schools in the Block 2a pool met the criteria to be considered as matches in Block 2b, we conducted the matching for the two blocks sequentially, with Block 2a first. Potential comparison schools that were not selected for Block 2a were added to the matching pool for Block 2b, which then consisted of 1,444 schools. For both Subsample 2 blocks, we matched on their average 2020 AP exam performance across all AP subject areas[7] as proxy measures corresponding to the AP CSP outcomes (see Chapter 5), along with enrollment size in grades 9 through 12.

For Block 2a, we experimented with selecting up to seven comparison schools as the number of comparison schools to be matched with each Lone Star school and decided on six as the optimal number (balancing strong baseline equivalence results with statistical power), resulting in a total of 84 comparison schools, 52 of which were in Texas (see top row of Exhibit 7.2).

For Block 2b, we initially attempted to use *kmatch md*, as usual, to select comparison schools. The 1:1 option yielded acceptable baseline equivalence on all four measures, but we deemed a total sample size of 12 schools (six Lone Star and six comparison) to be too small. All of the options with a higher comparison-to-Lone-Star school ratio failed to achieve baseline equivalence on the first measure (i.e., average percentage, across all 38 AP subject area exams, of 2020 exam takers earning a 3, 4, or 5). We then increased the allowable ratio to as many as 15 possible comparison schools per Lone Star school but still could not achieve baseline equivalence on the first measure. At that point we gave up on Mahalanobis distance matching to identify the Block 2b comparison schools for the Year 3 analysis and turned to two possible alternatives. One was to use *kmatch* with propensity score matching (option *ps*) instead of distance matching; the other was to use a hand-matching tool (Holtzman, 2024).[8] Both of these methods yielded sets of comparison schools with baseline equivalence on all four measures, but the baseline SMDs achieved with the hand-matching tool were notably lower than those achieved with *kmatch ps*. Thus, we went with the results from the hand-matching process,

[6] See footnote 15 in Chapter 5 for an explanation for why we did not limit the comparison schools in this block to those that, like the Lone Star schools in the block, had no AP CSP exam takers in 2023. In the matching process, we purposely remained blind to whether or not the potential comparison schools had AP CSP exam takers in 2023.
[7] See footnote 16 in Chapter 5 for additional detail about why we included all AP subject areas.
[8] This tool, which currently allows for matching on up to four continuous variables, allows the user to try out different matching solutions by selecting different potential comparison schools (and, if necessary, different intervention schools), with instantaneous feedback on how the baseline SMD (Hedges' *g)* for each measure changes—and whether it meets WWC criteria for baseline equivalence—when different schools are selected. The means (as well as the standard deviations) of each variable for each set of selected schools (intervention and comparison) are also displayed to help the user determine which schools to select or deselect to make the two groups more similar. The user can decide how many comparison schools to select; there need not be an exact ratio of intervention schools to comparison schools.

which yielded a set of 36 comparison schools, 30 of which were in Texas (see second row of Exhibit 7.2).[9]

Exhibit 7.2. States of the Year 3, Subsample 2 Comparison Schools

Block	Texas	Georgia	Massachusetts	New York	Wisconsin	Total
2a	52	4	12	13	3	**84**
2b	30	5	1	0	0	**36**
Subsample 2 Total	**82**	**9**	**13**	**13**	**3**	**120**

The within-block baseline equivalence statistics for each of the Subsample 2 blocks, which were used to select the comparison schools within each block, are provided in Appendix Exhibit F-2. The combined-block baseline equivalence statistics, calculated as described in Chapter 5, are presented in Exhibit 7.3. As shown in the last column, the Hedges' *g* for each baseline measure was less than (an absolute value of) 0.25 standard deviations, which meets WWC criteria provided that the baseline measure is included as a covariate in the impact analysis for its corresponding outcome measure (which it was).

[9] As noted in an earlier footnote, we purposely remained blind, during the matching, to whether the schools in the Block 2b comparison pool had AP CSP exam takers in 2023. It turned out that of the 36 selected schools, three had AP CSP exam takers in 2023.

Exhibit 7.3. Baseline Statistics for the Year 3, Subsample 2 Analysis Sample

Measure	Lone Star schools (n = 20)	Comparison schools (n = 120)	Regression coefficient (β_{Trt})	p value	Hedges' g
Average percentage (across all 38 AP subject area exams) of 2020 exam takers earning a 3, 4, or 5	34.59% (15.53)	35.96% (16.72)	-1.37%	0.718	-0.082
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 3, 4, or 5 in 2020	0.29% (0.26)	0.29% (0.27)	0.00%	0.944	-0.017
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 1 or 2 in 2020	0.47% (0.37)	0.51% (0.57)	-0.04%	0.757	-0.072
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 taking a 2020 AP exam	0.76% (0.55)	0.80% (0.74)	-0.04%	0.799	-0.061

Note. In the first two columns, each cell contains the unadjusted mean and (in parentheses) standard deviation of the baseline measure. In the third column, the regression coefficient is the coefficient on the Lone Star indicator from the regression of the baseline measure on the Lone Star indicator and matching block indicator (i.e., the block-adjusted baseline mean difference). The Hedges' g value results from the inclusion of the regression coefficient as the numerator (i.e., the block-adjusted mean difference) in the Hedges' g formula.

In summary, for both of the two subsamples, we succeeded in selecting a well-matched set of comparison schools—equivalent to the Lone Star schools at baseline on the outcomes to be examined—for the 2023 analysis of program impacts on AP CSP exam performance and participation.

Statistical Estimation Strategy

The statistical estimation strategy we used to estimate the effect of the Lone Star program on the 2023 AP CSP exam performance and participation outcomes (i.e., impact after three years of program implementation) was the same as that used for the Year 1 analysis (see Chapter 5). Specifically, we began by running a separate ordinary least squares (OLS) regression for each of the four outcomes in each of the two subsamples, with a treatment indicator as the key predictor of interest, a baseline measure of the outcome as a key covariate to adjust for minor baseline differences, a block indicator if needed, and school demographic characteristics as additional covariates.[10] The school demographic covariates included in the regression models were the same as those in the Year 1 analysis except that they (like the outcome measures)

[10] See Chapter 5 for detail about why we conducted a separate regression for each subsample and how the models for the two subsamples differed from one another.

were from the 2022–23 school year;[11] these data were drawn from the Common Core of Data,[12] except for the economically disadvantaged variable for Massachusetts schools, which came from state data.[13]

Appendix Exhibits F-3 and F-4 provide descriptive statistics on all of the covariate measures for the Subsample 1 Lone Star and comparison schools, and Appendix Exhibits F-5 and F-6 provide the covariate measure descriptives for the Lone Star and comparison schools in Subsample 2. As was the case in the Year 1 and Year 2 samples of schools (see Chapters 5 and 6), there were some notable demographic differences between the Lone Star schools and the selected comparison schools, as indicated by the effect sizes shown in the rightmost column of each table. These differences were generally similar to those observed in the previous years:

- In both Subsample 1 and Subsample 2, Lone Star schools were much more likely than the comparison schools to be charter schools.
- In Subsample 1, Lone Star schools, on average, had much higher proportions of Hispanic students, slightly lower proportions of female students, and higher percentages of economically disadvantaged students than comparison schools. They were also much more likely to be located in cities and less likely to be located in suburbs or in towns or rural areas.
- In Subsample 2, Lone Star schools had notably lower percentages of female students, on average, than comparison schools. They were also notably more likely to be in cities and much less likely to be in rural areas or towns.

As previously mentioned, our statistical models controlled for these and other school characteristics.

After conducting the regressions, we calculated the effect size of the estimated intervention effect for each of the four measures in each of the two subsamples, using the *adjusted* mean difference in the *outcome* year, namely the β_{Trt} coefficient from the regression, as the numerator in the Hedges' *g* formula. We then conducted a meta-analysis combining the

[11] See footnote 19 in Chapter 5 for an explanation for why we used the outcome year (in this chapter, 2022–23) rather than the 2019–20 baseline year as the year for all school demographic characteristics.
[12] Common Core of Data: https://nces.ed.gov/ccd/elsi/tableGenerator.aspx
[13] For schools in all states except Massachusetts, the economically disadvantaged covariate was the school percentage of students in the free and reduced-price lunch (FRL) program based on counts provided in the Common Core of Data (CCD). However, FRL data were not available in CCD for Massachusetts schools in any of our analysis years, so we instead used percentages of "low income" students, based on counts in data files at https://profiles.doe.mass.edu/statereport/selectedpopulations.aspx. Because of the difference between states on this variable, we standardized the values within state. One New York school and three Texas schools were missing economically disadvantaged data; so as not to exclude these schools, we imputed the missing values as zero (i.e., the within-state standardized mean) and included an indicator of missingness in the model. No other variables had missing data.

estimates from the two subsamples to yield a single overall estimate for each outcome. (See Chapter 5 for further detail on how the meta-analysis was conducted.)

Results

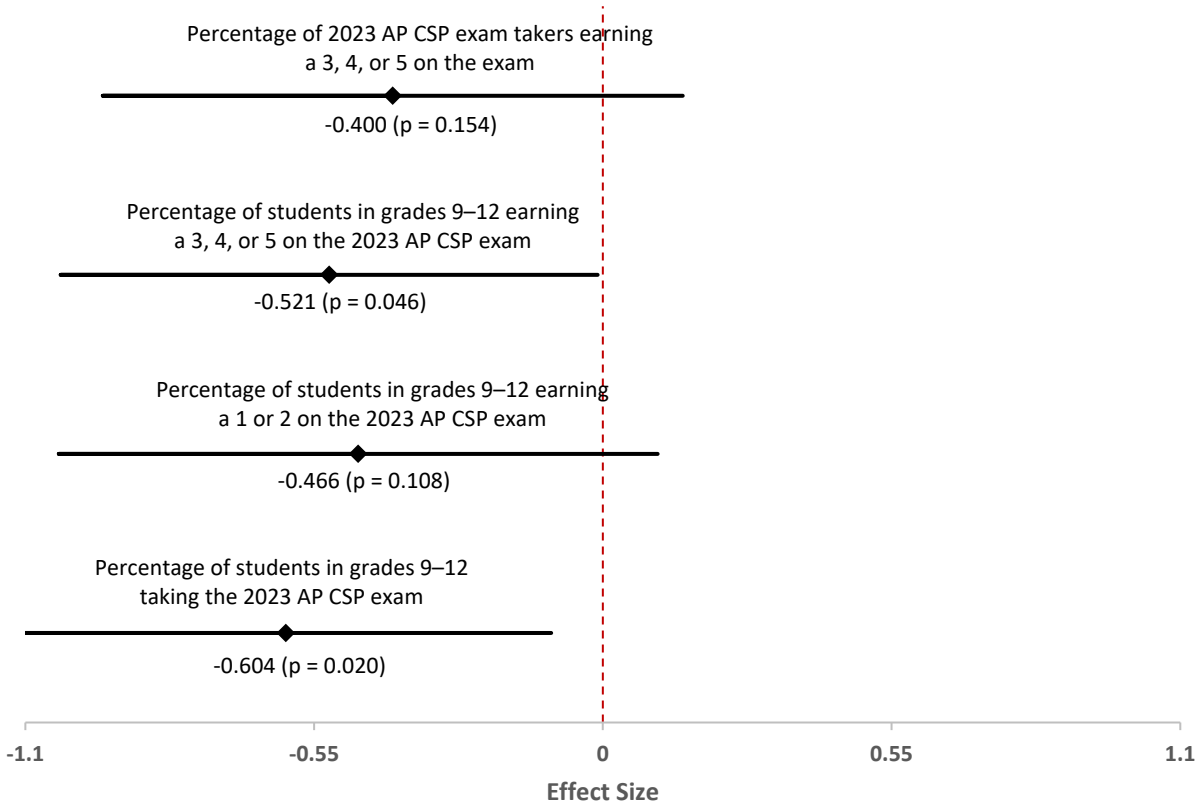
In this section we first present the results separately for each of the two subsamples and then present the meta-analysis (combined) results.

Subsample 1 and 2 (OLS-Based) Results

Exhibits 7.4 and 7.5 show the estimated effect sizes for the third-year effects of the Lone Star AP CSP program on the four outcome measures for Subsample 1 and Subsample 2, respectively. Exhibit 7.6 provides additional details, including the unadjusted (i.e., descriptive) means and standard deviations of each outcome measure, the regression-adjusted mean difference (i.e., the coefficient on the Lone Star indicator from the OLS regression), the *p* value of the coefficient from the regression, and the calculated Hedges' *g* effect size (identical to the point estimates shown in Exhibits 7.4 and 7.5). The full regression results are provided in the appendix (Exhibits F-7 and F-8 for each subsample, respectively).

The Subsample 1 results are quite similar to those observed in Year 2. For schools in Subsample 1 (i.e., those that had AP CSP exam takers in the baseline year), the Lone Star AP CSP program again appears to have had a statistically significant effect on two outcomes, and the same two as in the Year 2 analyses: percentage of students in grades 9–12 earning a 3, 4, or 5 on the AP CSP exam and percentage of students in grades 9–12 taking the exam (Exhibit 7.4). As in Year 2, both effects are negative and moderate in size, about one half of a standard deviation. Also as in Year 2, the estimates for the other two outcomes—percentage of 2023 AP CSP exam takers earning a 3, 4, or 5 on the exam and percentage of students in grades 9–12 earning a 1 or 2 on the 2023 exam—are negative but not statistically significant.

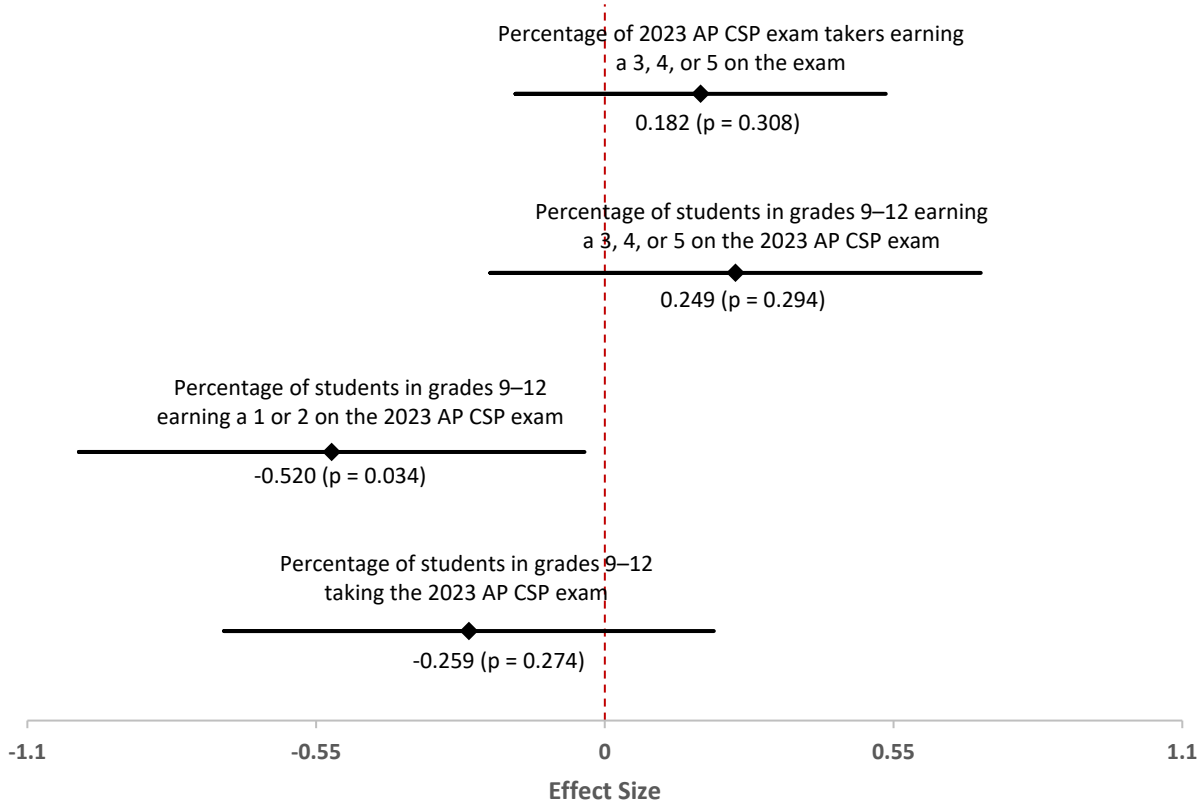
Exhibit 7.4. Effect Size Chart for the Year 3, Subsample 1 Analysis



Note. *N* = 108 (12 Lone Star schools and 96 comparison schools). Effect size point estimates are represented by the diamonds; the horizontal lines show 95% confidence intervals. Confidence intervals that do not intersect with an effect size of 0 (represented by the dotted red line) indicate a statistically significant ($p < .05$) effect.

Unlike the Subsample 1 results, the Subsample 2 results are quite different from those in Year 2, and, correspondingly, they are also quite different from the Subsample 1 results. For Subsample 2, the effects for the three performance measures were all in the desired direction: a positive effect for the percentage of 2023 AP CSP exam takers earning a qualifying score on the exam, a positive effect for the percentage of students in grades 9–12 earning a qualifying score, and a negative effect for the percentage of students in grades 9–12 earning a nonqualifying score (Exhibit 7.5). Moreover, the last of these was statistically significant and of moderate magnitude. For the participation measure—the percentage of students in grades 9–12 taking the 2023 AP CSP exam—the Year 3 effect in Subsample 2 was negative but not statistically significant. Thus it appears that although (relative to the comparison schools) fewer students in the Lone Star Subsample 2 schools took the AP CSP exam in Year 3, more of them performed well on it, and more notably, significantly fewer of them performed poorly on it.

Exhibit 7.5. Effect Size Chart for the Year 3, Subsample 2 Analysis



Note. *N* = 140 (20 Lone Star schools and 120 comparison schools). Effect size point estimates are represented by the diamonds; the horizontal lines show 95% confidence intervals. Confidence intervals that do not intersect with an effect size of 0 (represented by the dotted red line) indicate a statistically significant ($p < .05$) effect.

Exhibit 7.6. Year 3 Outcome Statistics for Both Subsamples

Measure	Lone Star schools (unadj.)	Comparison schools (unadj.)	Adj. mean difference (β_{Trt})	*p* value	Hedges' *g*
Subsample 1 (12 Lone Star schools and 96 comparison schools)					
Percentage of 2023 AP CSP exam takers earning a 3, 4, or 5 on the exam	49.28% (24.97)	66.83% (23.50)	-9.54% (6.63)	0.154	-0.400
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2023 AP CSP exam	1.25% (1.55)	1.53% (1.42)	-0.75% (0.37)	0.046	-0.521
Percentage of students in grades 9–12 earning a 1 or 2 on the 2023 AP CSP exam	1.02% (0.73)	0.82% (1.12)	-0.51% (0.31)	0.108	-0.466
Percentage of students in grades 9–12 taking the 2023 AP CSP exam	2.27% (1.74)	2.35% (2.16)	-1.29% (0.54)	0.020	-0.604
Subsample 2 (20 Lone Star schools and 120 comparison schools)					
Percentage of 2023 AP CSP exam takers earning a 3, 4, or 5 on the exam	25.61% (32.42)	29.06% (29.23)	5.45% (5.32)	0.308	0.182
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2023 AP CSP exam	1.15% (2.46)	0.62% (1.03)	0.33% (0.31)	0.294	0.249
Percentage of students in grades 9–12 earning a 1 or 2 on the 2023 AP CSP exam	1.01% (1.31)	0.97% (2.04)	-1.02% (0.48)	0.034	-0.520
Percentage of students in grades 9–12 taking the 2023 AP CSP exam	2.16% (2.91)	1.60% (2.65)	-0.70% (0.64)	0.274	-0.259

Note. In the first two columns, each cell contains the unadjusted mean and (in parentheses) standard deviation of the outcome measure. The third column contains the regression coefficient on the Lone Star indicator from the impact analysis model (i.e., the adjusted mean difference) and (in parentheses) the standard error of the coefficient; the *p* value associated with the coefficient is in the fourth column. The Hedges' *g* value in the final column results from the inclusion of the regression coefficient as the numerator (i.e., the mean difference) in the Hedges' *g* formula.

The significant negative influence of schools' percentages of female students on the participation outcome, as observed in the Year 2 analysis and discussed in Chapter 6, was less pronounced in Year 3, showing up (as statistically significant) only for Subsample 1. And unlike in Year 2, even descriptively, the comparison schools in Subsample 1 had a higher percentage of

students taking the AP CSP exam than did the Lone Star schools, although this unadjusted difference was very small (much smaller than the *adjusted* difference resulting from the covariate-controlled regression analysis).

In both subsamples in the Year 3 analysis, being a charter school significantly and positively predicted the AP CSP participation outcome. Because the Lone Star schools were more likely to be charter schools than the comparison schools, the inclusion of the charter school indicator as a covariate likely reduces the treatment estimate.

Meta-Analysis Results

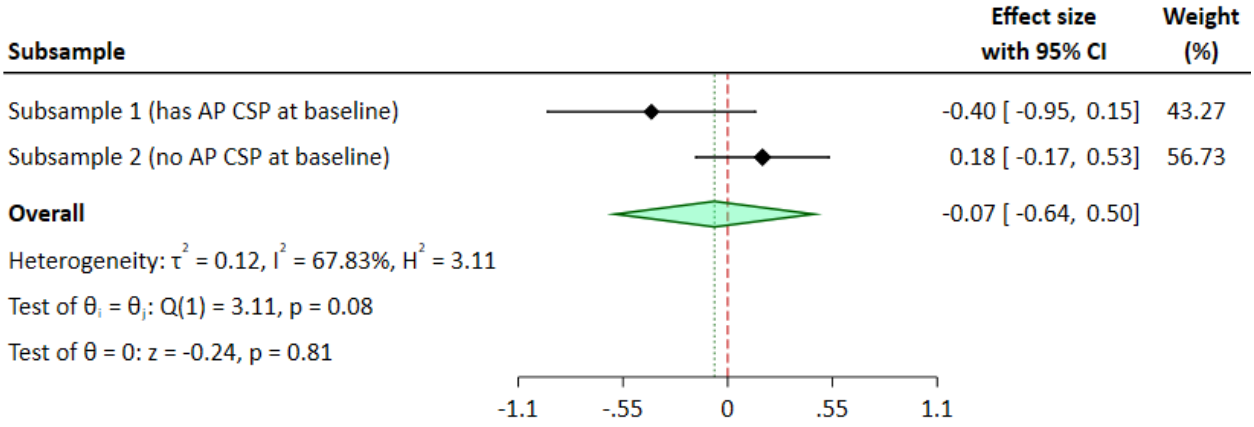
Exhibits 7.7 through 7.10 display the Year 3 meta-analysis results for each of the four outcome measures, respectively. As the exhibits show, there was a statistically significant, moderately sized negative overall effect for two of the 2023 outcomes: percentage of students in grades 9–12 earning a nonqualifying score on the AP CSP exam (Exhibit 7.9) and percentage of students in grades 9–12 taking the AP CSP exam (Exhibit 7.10). These were the two outcomes for which the estimates in both individual subsample analyses were negative, although not significantly so in both. (The estimate for the former was significant only in the Subsample 2 analysis, while the estimate for the latter was significant only in the Subsample 1 analysis.) For each of these two outcomes, the meta-analysis showed no significant variation between the two samples or significant difference between the two estimates, so when the estimates are combined in the meta-analysis (which allows for greater statistical power), the overall effect is significant.

The significant negative estimate for the participation outcome is not in the desired direction—that is, it appears that the Lone Star AP CSP program had a negative effect on schools' proportions of students taking the AP CSP exam, relative to matched comparison schools and controlling for baseline participation levels and school demographics. On the other hand, the significant negative effect of the estimate for the performance outcome—the percentage of students in the school earning a *non*qualifying score on the AP CSP exam—is in the desired direction; the program appears to have contributed to fewer students in the school doing poorly on the exam, all else held constant.

The overall estimates for the other two outcomes—percentage of AP CSP exam takers earning a qualifying score (Exhibit 7.7) and percentage of students in grades 9–12 earning a qualifying score on the AP CSP exam (Exhibit 7.8)—were close to zero and not significant. These were the two outcomes where the two subsamples' results did not resemble one another, with Subsample 1 showing negative effects and Subsample 2 showing positive effects. With the two samples' estimates averaging one another out, it is not surprising that the overall effect for each outcome is close to zero. Indeed, unlike the other two outcomes in the Year 3 analysis, and unlike *any* of the four outcomes in either prior year's analysis, the meta-analysis for each of

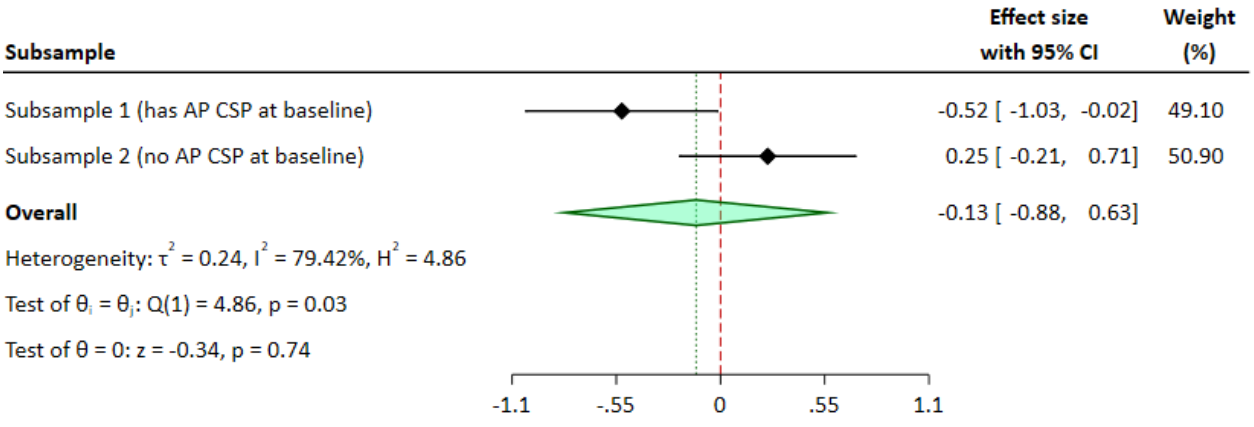
these two Year 3 outcomes shows that there was notable variation between Subsample 1 and Subsample 2—a borderline significant difference ($p < .10$) between the two subsamples for the first outcome and a significant difference ($p < .05$) between the two subsamples for the second.

Exhibit 7.7. Meta-Analysis Plot for Percentage of 2023 AP CSP Exam Takers Earning a 3, 4, or 5 on the Exam



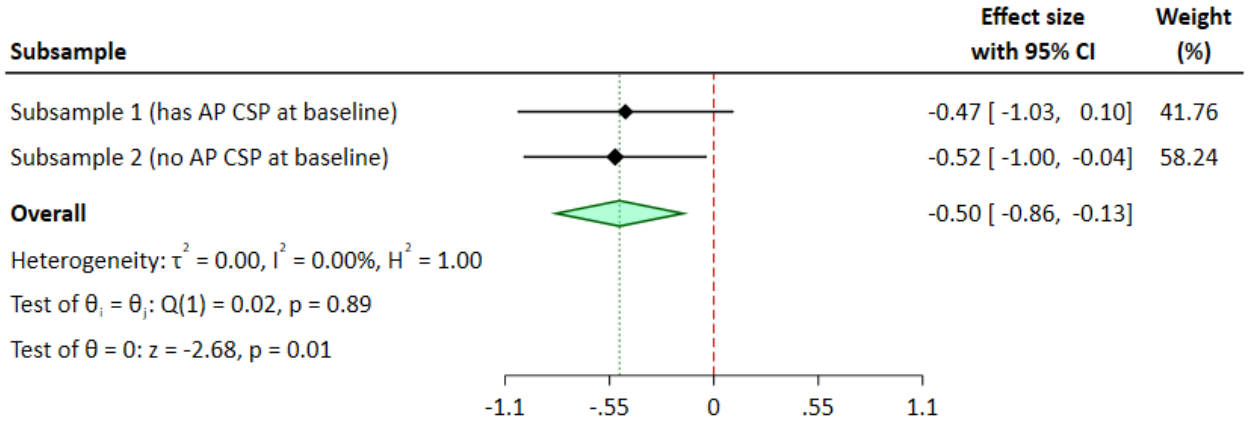
Note. See box in Chapter 5 titled "How to Read the Meta-Analysis Figures."

Exhibit 7.8. Meta-Analysis Plot for Percentage of Students in Grades 9–12 Earning a 3, 4, or 5 on the 2023 AP CSP Exam



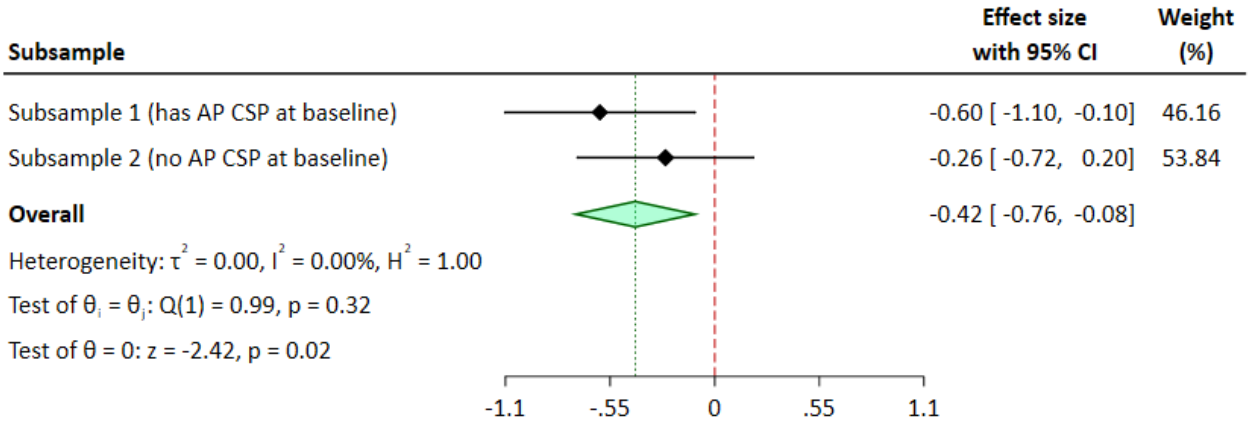
Note. See box in Chapter 5 titled "How to Read the Meta-Analysis Figures."

Exhibit 7.9. Meta-Analysis Plot for Percentage of Students in Grades 9–12 Earning a 1 or 2 on the 2023 AP CSP Exam



Note. See box in Chapter 5 titled "How to Read the Meta-Analysis Figures." For test of θ, $p = 0.0074$.

Exhibit 7.10. Meta-Analysis Plot for Percentage of Students in Grades 9–12 Taking the 2023 AP CSP Exam



Note. See box in Chapter 5 titled "How to Read the Meta-Analysis Figures."

Summary and Discussion

In summary, we found that after three years of implementation, the Lone Star AP CSP program appears to have had a statistically significant negative effect on two outcomes: the percentages of students in grades 9–12 that took the AP CSP exam and the percentages of students in grades 9–12 that earned a nonqualifying score (i.e., a 1 or a 2) on the AP CSP exam. As relates to the Lone Star program's theory of action, the latter is in the desired direction (fewer students earning a nonqualifying score is a good thing), but the former is not.

The negative participation effect was also found in Year 2, so much of the discussion about that finding in the Chapter 6 discussion applies here as well. As noted in Chapter 2, the school supports component of the Lone Star AP CSP program, which according to the logic model was theorized to be a key driver of increased rates of participation in AP CSP, did not meet fidelity-of-implementation criteria in either Year 2 (which could have affected students' enrollment in AP CSP in Year 3) or Year 3. The teacher supports component, which could also have influenced whether AP CSP students took the exam, also was not met in Year 3.

The significant negative effect of the program on the percentage of students earning a *non*qualifying score is likely related to the negative participation effect: in the average Lone Star school (relative to the average comparison school), fewer students in the school taking the AP CSP exam translated to fewer students in the school doing poorly on the exam. In Subsample 1, the strong negative participation effect was also reflected in lower percentages of students in the school who did well (i.e., earned a qualifying score) on the exam. But this was not the case in Subsample 2, where the negative participation effect was not as strong as in Subsample 1 (and not, on its own, statistically significant). In fact, in Subsample 2, there were positive, but not statistically significant, effects on the percentages of students who earned a qualifying score on the exam, both as a percentage of the entire grades 9–12 enrollment and as a percentage of students who took the exam. This might suggest that in Year 3, the Subsample 2 Lone Star schools, despite being newer to AP CSP than the Subsample 1 Lone Star schools, were more effective in identifying students who were likely to succeed on the AP CSP exam, and were having only these students take the exam (and possibly also having only these students take the AP CSP course; the evaluation lacks course enrollment data to make this distinction). NMSI's theory holds that *more* students, not fewer students, are capable of succeeding in AP CSP (and on the AP CSP exam) than are typically enrolled, particularly when strong supports are provided, but it is possible that this is not the case, or that the supports were insufficient for schools to decide that more students were likely to succeed.

Addendum to Chapters 5–7. Impact on Student AP CSP Performance and Participation Using a Constant, All-Texas Comparison Group for 12 Subsample 1 Lone Star Schools (Robustness Check)

As indicated in Chapters 5, 6, and 7, we matched Lone Star schools with similar comparison schools separately for each of the three years' analysis. We conducted separate matching processes each year because not all schools had AP CSP exam takers every year, presumably at least in large part because they did not offer AP CSP every year. This was true even for the Lone Star AP CSP schools, particularly those in Subsample 2. As shown by the box in Chapter 7 about Lone Star schools' cross-year patterns of AP CSP exam takers, the 37 Lone Star schools collectively displayed 14 different patterns of AP CSP exam taking across the four years covered by the evaluation (i.e., the baseline year plus the three outcome years). Non-Lone-Star schools in each year's pool of potential comparison schools likely were similarly varied in their cross-year patterns of AP CSP exam takers. Therefore, a non-Lone-Star school selected as a matched comparison school in any given outcome year's analysis could not necessarily be counted upon to serve as a comparison school in a different year. Thus, for each of the three outcome years, we conducted a separate matching process to identify similar comparison schools for the Lone Star schools included in that year's analysis. Although this was necessary, it also meant that if the study results differed across the three years—as turned out to be the case—we could not know for certain whether the differences were because of true year-to-year differences in program effectiveness or because of the different samples of schools included each year.

Among the 37 Lone Star schools, the modal pattern for AP CSP exam taking—that is, the pattern displayed by the largest number of schools—was having (at least five) AP CSP exam takers in all four years. Twelve Lone Star schools—by definition all in Subsample 1 (given AP CSP exam takers in the baseline year), and in fact including all but *one* of the Subsample 1 schools—had this pattern. Upon the conclusion of the main analyses reported in Chapters 5, 6, and 7, we were curious to see whether it would be possible to select a suitably large *constant* set of well-matched comparison schools—that is, a single set of comparison schools that would be the same in all three outcome years—for these 12 Lone Star schools. We decided to try it. Moreover, given the questions raised by the Year 2 results in particular, we decided to also see if it would be possible to select a suitably large constant set of comparison schools that were *all located in Texas*.

There were 166 potential comparison schools that, like the 12 Lone Star schools, had (unsuppressed) AP CSP exam data in the baseline year and in all three outcome years and were

in Texas. Again using the Stata module *kmatch* (Jann, 2017) with multivariate-distance matching (option *md,* with the default Mahalanobis distance matching) and specifying nearest-neighbor matching (*nn*) without replacement (*wor*), matching on the same variables as we had for the main-analysis matching for Subsample 1, we experimented with selecting 1 to 10 comparison schools as the number of schools to be matched with each of the 12 Lone Star schools. The match solutions with one, two, three, and seven comparison schools per Lone Star school all yielded mean differences of less than 0.25 standard deviations on all four outcomes at baseline as well as on grades 9–12 enrollment at baseline. Clearly the seven-school solution—yielding a total of 84 comparison schools—was better than the others in terms of statistical power. Exhibit ADD-1 displays the means, standard deviations, and standardized mean differences for the 12 Lone Star schools and the 84 comparison schools selected in this manner.[1]

Exhibit ADD-1. Baseline Statistics for Subsample 1 Constant All-Texas Sample

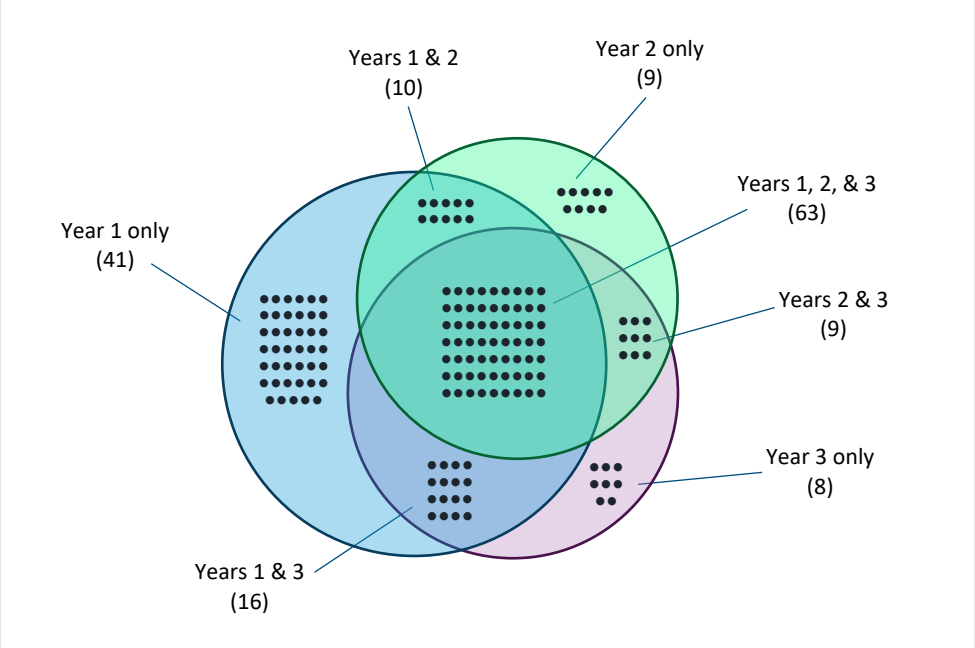
Measure	Lone Star schools (*n* = 12)	Comparison schools (*n* = 84)	*p* value	Hedges' *g*
Percentage of 2020 AP CSP exam takers earning a 3, 4, or 5 on the exam	68.66% (18.02)	67.55% (19.55)	0.854	0.057
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2020 AP CSP exam	1.14% (0.92)	1.34% (1.49)	0.663	-0.134
Percentage of students in grades 9–12 earning a 1 or 2 on the 2020 AP CSP exam	0.41% (0.27)	0.54% (0.59)	0.448	-0.233
Percentage of students in grades 9–12 taking the 2020 AP CSP exam	1.55% (1.08)	1.88% (1.94)	0.571	-0.174
Number of students enrolled in grades 9–12, 2019–20	1,644 (1,180)	1,875 (1,064)	0.488	-0.213

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from independent-samples *t* tests; they are provided for parallelism with the outcome analysis results. Hedges' *g* in the last column is calculated as explained in previous chapters.

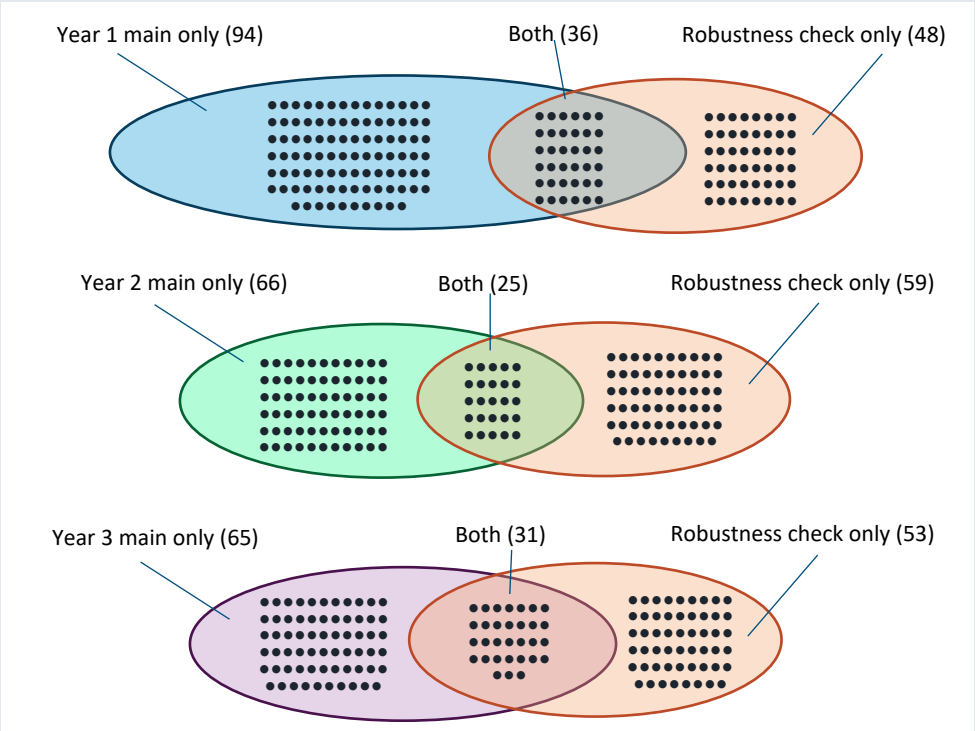
Therefore, it did indeed turn out to be possible to select a set of constant, all-Texas comparison schools that were equivalent at baseline on AP CSP performance and participation to the 12 Lone Star schools.

[1] It is worth noting that although these standardized mean differences, at less than 0.25 standard deviations, are acceptable by WWC standards (provided that the baseline measure is included as a covariate in the impact analysis for its corresponding outcome measure), they are not as low as the standardized mean differences we obtained for Subsample 1 in the main analyses (Chapters 5, 6, and 7), which were all less than 0.05 standard deviations. In other words, for the measures of the outcomes at baseline, the Subsample 1 comparison schools selected for the main analysis were more similar to the Lone Star schools than the alternate sample of comparison schools selected here.

OVERLAP AMONG THE THREE YEARS' SUBSAMPLE 1 COMPARISON SCHOOLS, MAIN ANALYSIS



OVERLAP BETWEEN EACH YEAR'S MAIN ANALYSIS SUBSAMPLE 1 COMPARISON SCHOOLS AND THE ROBUSTNESS CHECK SAMPLE COMPARISON SCHOOLS



Reanalysis

Given that we were able to identify a constant set of all-Texas comparison schools for the 12 Lone Star schools, the next logical question was whether the results would look different if we redid each year's analyses using this alternate set of comparison schools.

We decided in advance that we would consider the results of any reanalyses with an alternate set of comparison schools to be a "robustness check" only and would stand by the results presented in the preceding chapters as the main results. We had three reasons for this decision. First, we were not sure that we would have done any of this exploration if the main analyses had yielded results that were more favorable to the Lone Star program. Second, one Lone Star school that we were able to include in the main analyses for Years 1 and 2 would be omitted from the reanalyses for those two years because it did not have AP CSP exam takers in Year 3. And finally, we did not want the comparison school selection process for Subsample 1 schools to be fundamentally different from that used for the Subsample 2 schools.[2]

The analysis method was the same as that used for Subsample 1 in the main analysis, including controlling for school background characteristics. This remained important because the alternate set of comparison schools, despite now all being in Texas, still did not perfectly match the 12 Lone Star schools on background characteristics. As indicated by the effect sizes shown in the rightmost column of Appendix Exhibits G-1 and G-2, the key differences were as follows:

- Compared to the comparison schools, the Lone Star schools were still notably more likely to be charter schools, although not by quite as much as in the main analysis samples.
- Compared to the comparison schools, the Lone Star schools were slightly more likely than the comparison schools to be located in cities and slightly less likely to be located in suburbs (though, again, not by as much as in the main analysis samples).

[2] It would have been very difficult, and perhaps impossible, to reselect comparison schools for Subsample 2 in the same way we did for Subsample 1. Most notably, there were nine different patterns of AP CSP exam taking among the 21 Subsample 2 Lone Star schools included in at least one of the three years' outcome analyses, compared to just two patterns among the 13 Subsample 1 Lone Star schools. Even if we had limited the attempt to the eight Subsample 2 Lone Star schools that had at least five AP CSP exam takers in all three outcome years (i.e., Block 2a schools in all three years)—which would have been of questionable value given the omission of the other 13 Subsample 2 schools (compared to only a single school omitted for Subsample 1)—it might not have been possible to select a constant set of all-Texas comparison schools across all three outcome years. The pool of potential comparison schools for Block 2a was already quite small; in particular, there were only 91 potential comparison schools in Year 1 (only 63 of which were in Texas), and likely only some of those schools also had AP CSP exam takers in both Years 2 and 3. Finally, even with the comparison schools selected for the main analyses, the baseline SMDs were generally higher for Subsample 2 than for Subsample 1; whereas for Subsample 1 we had baseline SMDs below 0.05 standard deviations in each of the three years, the baseline SMDs for Block 2a tended to be higher (though still below the maximum allowed 0.25 standard deviations), and it would have been that much more difficult to achieve baseline equivalence with a smaller comparison pool (i.e., a pool limited to Texas schools that had AP CSP exam takers in all three years). Moreover, in terms of the all-Texas specification, there was less of a need in Subsample 2 than in Subsample 1. In each of the three years, more than two thirds of the selected Subsample 2 comparison schools were in Texas (coincidentally, 68% each year), compared to much lower percentages in Subsample 1 (41% in Year 1, 37% in Year 2, and 39% in Year 3).

- Compared to the comparison schools, the Lone Star schools had notably higher percentages of Hispanic students (though not by nearly as much as in the main analysis samples, which is not surprising given all of the schools now being in Texas).
- Compared to the comparison schools, the Lone Star schools had notably higher percentages of students who were economically disadvantaged. These differences were actually slightly larger than in the main analysis samples.
- Compared to the comparison schools, the Lone Star schools had very slightly lower percentages of female students. These differences were similar to those in the main analysis samples for Years 1 and 3, but notably smaller than the difference in the main analysis sample for Year 2.

Therefore, although for some school characteristics the alternate set of comparison schools offered greater similarity to the Lone Star schools than in the main analysis sample, for others it did not.

Results

In this section we focus on comparing the results from the robustness check with those from the main analyses presented in Chapters 5, 6, and 7; displays of results from the robustness check analyses that parallel those for the main analysis are presented in the appendix. (Specifically, Appendix Exhibit G-3, G-4, and G-5 include effect size charts for the robustness check analyses for each of the three years, respectively; these exhibits also include, for the sake of easy comparison, the corresponding effect size charts from the main analysis. Appendix Exhibit G-6 provides the outcome statistics for all three years' robustness check analyses, and Exhibits G-7, G-8, and G-9 provide the full model regression results for each of the three years' robustness check analyses, respectively.)

Exhibit ADD-2. Comparison Between Main Analysis for Subsample 1 and Robustness Check: Hedges' *g*, Standard Error of Hedges' *g*, and Regression *p* Value

a. 2021 (Year 1)

Outcome	Main analysis – Ch. 5 (13 Lone Star & 130 comparison schools)	Robustness check (12 Lone Star & 84 comparison schools)
Percentage of 2021 AP CSP exam takers earning a 3, 4, or 5 on the exam	-0.004 (0.239) 0.986	0.178 (0.236) 0.452
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2021 AP CSP exam	0.116 (0.242) 0.632	0.070 (0.212) 0.741
Percentage of students in grades 9–12 earning a 1 or 2 on the 2021 AP CSP exam	0.156 (0.199) 0.435	-0.410* (0.196) 0.039
Percentage of students in grades 9–12 taking the 2021 AP CSP exam	0.159 (0.216) 0.463	-0.377* (0.189) 0.049

***p < .001, **p < .01, *p < .05, †p < .10

b. 2022 (Year 2)

Outcome	Main analysis – Ch. 6 (13 Lone Star & 91 comparison schools)	Robustness check (12 Lone Star & 84 comparison schools)
Percentage of 2022 AP CSP exam takers earning a 3, 4, or 5 on the exam	-0.352 (0.259) 0.177	-0.486* (0.216) 0.027
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2022 AP CSP exam	-0.526* (0.218) 0.018	-0.429* (0.211) 0.045
Percentage of students in grades 9–12 earning a 1 or 2 on the 2022 AP CSP exam	-0.270 (0.192) 0.165	-0.317 (0.231) 0.175
Percentage of students in grades 9–12 taking the 2022 AP CSP exam	-0.480** (0.166) 0.005	-0.380† (0.215) 0.081

***p < .001, **p < .01, *p < .05, †p < .10

c. 2023 (Year 3)

Outcome	Main analysis – Ch. 7 (12 Lone Star & 96 comparison schools)	Robustness check (12 Lone Star & 84 comparison schools)
Percentage of 2023 AP CSP exam takers earning a 3, 4, or 5 on the exam	-0.400 (0.278) 0.154	0.115 (0.207) 0.579
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2023 AP CSP exam	-0.521* (0.258) 0.046	-0.044 (0.217) 0.838
Percentage of students in grades 9–12 earning a 1 or 2 on the 2023 AP CSP exam	-0.466 (0.287) 0.108	-0.431† (0.224) 0.058
Percentage of students in grades 9–12 taking the 2023 AP CSP exam	-0.604* (0.255) 0.020	-0.493* (0.211) 0.022

***$p < .001$, **$p < .01$, *$p < .05$, †$p < .10$

Note. Positive effect sizes indicate effects in the desired direction according to the theory of action, except for each panel's third row outcome (percentage of students in grades 9–12 earning a 1 or 2 on the AP CSP exam), where negative effect sizes are in the desired direction.

Exhibit ADD-2 shows that in no case—for any of the four outcomes in any of the three years—did an estimate that was significantly negative in the main analysis become significantly positive in the robustness check, nor did any estimate go from significantly positive to significantly negative. Nevertheless, there were some differences between the main Subsample 1 analysis results and the robustness check results:

- In Year 1, the estimates for three of the four outcomes switched sign, and two of them switched from nonsignificant to significant:
  - The estimate for the percentage of students in grades 9–12 earning a 1 or 2 on the 2021 AP CSP exam went from nonsignificant and positive to significant ($p < .05$) and negative, which is the desired direction for this outcome.
  - The estimate for the percentage of students in grades 9–12 taking the exam went from nonsignificant and positive to significant ($p < .05$) and negative, which is not in the desired direction.
- In Year 2, the estimated effects for all four outcomes in the robustness check were negative, as was also the case in the main analysis for Subsample 1. But the effect for the percentage of 2022 AP CSP exam takers earning a 3, 4, or 5 on the exam went from nonsignificant to significant ($p < .05$), while the effect for the percentage of students in

grades 9–12 taking the exam went from significant at $p < .01$ to only borderline significant ($p < .10$).

- In Year 3, the estimates for three of the four outcomes had a difference of some type:
  - The estimate for the percentage of 2023 AP CSP exam takers earning a 3, 4, or 5 on the exam switched sign but was not significant in either analysis (meaning that neither estimate could be assumed, with a high degree of confidence, to be different from zero).
  - The estimate for the percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2023 AP CSP exam went from significant ($p < .05$) and negative to nonsignificant (but still negative).
  - The estimate for the percentage of students in grades 9–12 earning a 1 or 2 on the 2023 AP CSP exam went from nonsignificant to borderline significant; both estimates were negative (which is in the desired direction for this outcome) and of similar magnitude (about 0.45 standard deviations).
  - The estimate for the percentage of students in grades 9–12 taking the 2023 AP CSP exam remained negative and significant at $p < .05$.

Despite this "wiggling around" of some of the estimates, the results yielded by the robustness check do not paint a wildly different picture of Lone Star AP CSP program effectiveness than the results from the main Subsample 1 analyses. This was particularly true for the key performance outcome (percentage of 2023 AP CSP exam takers earning a 3, 4, or 5 on the exam) and the participation outcome; the robustness check did not yield a more favorable impression of the program for either of these outcomes. For the supplementary outcomes for performance—particularly the percentage of students in grades 9–12 earning a 1 or 2 on the AP CSP exam—the robustness check results were somewhat more favorable than the main analyses in Year 1 and Year 3, but, as we have stated in prior chapters, these findings are probably closely linked to the accompanying findings of lower AP CSP participation rates.

The differences in results between the main Subsample 1 analyses and the robustness check analyses could be because of the one Lone Star school omitted from the robustness check analyses (except in Year 3, when the sample of Lone Star schools in the main analysis was exactly the same as that in the robustness check), the different set of comparison schools, or both. (See the box earlier in this chapter for a visual display of the amount of overlap between each year's main analysis Subsample 1 comparison schools and the set of 84 comparison schools included in each year's robustness check.)

As an additional comparison between the main Subsample 1 analysis and the robustness check analysis, we also tested the degree of heterogeneity between the two effect estimates for each

outcome in each year.[3] In only one of the 12 cases (three years times four outcomes) did we find significant heterogeneity: for the percentage of students in grades 9–12 earning a 1 or 2 on the 2021 AP CSP exam (p = 0.043). There also was borderline significant heterogeneity for the percentage of students in grades 9–12 taking the 2021 AP CSP exam (p = 0.061). The lack of significant heterogeneity in nearly all cases is further evidence that the results yielded by the robustness check were generally not that different than the results yielded by the main analysis.

Nevertheless, to be maximally thorough, we decided to redo the meta-analysis for each outcome in each year, keeping the Subsample 2 results as they were, but replacing the main Subsample 1 results with the robustness check results. The forest plots from these meta-analyses are in Appendix Exhibits G-10, G-11, and G-12. Exhibit ADD-3 presents a summary comparing the meta-analysis overall effect sizes from the main analysis with those that instead used the robustness check results for Subsample 1.

[3] We did this by including the two effect size estimates and their associated standard errors (for each outcome in each year) in a meta-analysis, but only to test for heterogeneity. (It would not have been appropriate to consider the overall effects from the meta-analysis given the overlap, especially in Lone Star schools, between the two samples and the fact that the main analysis and the robustness check were both being used for the same purpose.)

Exhibit ADD-3. Comparison of Meta-Analysis Results (Overall Effect Size θ) from Main Analyses and Robustness Check Analyses

Outcome	θ using main analysis Subsample 1	θ using robustness check Subsample 1
Year 1		
Percentage of AP CSP exam takers earning a 3, 4, or 5 on the exam	-0.11	-0.02
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the AP CSP exam	0.13	0.10
Percentage of students in grades 9–12 earning a 1 or 2 on the AP CSP exam	0.14	-0.15
Percentage of students in grades 9–12 taking the AP CSP exam	0.10	-0.19
Year 2		
Percentage of AP CSP exam takers earning a 3, 4, or 5 on the exam	-0.27†	-0.33*
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the AP CSP exam	-0.49**	-0.44**
Percentage of students in grades 9–12 earning a 1 or 2 on the AP CSP exam	-0.23†	-0.24†
Percentage of students in grades 9–12 taking the AP CSP exam	-0.40**	-0.33*
Year 3		
Percentage of AP CSP exam takers earning a 3, 4, or 5 on the exam	-0.07	0.15
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the AP CSP exam	-0.13[a]	0.09
Percentage of students in grades 9–12 earning a 1 or 2 on the AP CSP exam	-0.50**	-0.47**
Percentage of students in grades 9–12 taking the AP CSP exam	-0.42*	-0.39*

***p < .001, **p < .01, *p < .05, †p < .10

[a] indicates significant (p < .05) heterogeneity between the two subsamples (Subsample 1 and Subsample 2).

Note. Effect sizes are from meta-analyses combining two subsamples of schools. Positive effect sizes indicate effects in the desired direction according to the theory of action, except for each panel's third row outcome (percentage of students in grades 9–12 earning a 1 or 2 on the AP CSP exam), where negative effect sizes are in the desired direction.

As shown by the table, there were no major differences, in terms of overall Lone Star AP CSP program effects, between the main meta-analysis results and the robustness check meta-analysis results. The Year 1 effects for all four outcome measures were all still nonsignificant and relatively small in magnitude, although two of them did change sign. The Year 2 effects remained mostly the same, the only notable difference being that the negative effect for the percentage of AP CSP exam takers earning a 3, 4, or 5 on the exam increased slightly in magnitude and went from being borderline significant in the main analysis to significant in the robustness check. (In addition, the negative effect for the percentage of students in grades 9–12 taking the AP CSP exam decreased slightly in magnitude and by one significance level, from $p < .01$ to $p < .05$.) The Year 3 effects all remained the same, in terms of statistical significance levels, and, for the two outcomes that had statistically significant effects, in sign and magnitude.

Summary

In short, we generally did not obtain different estimates of program effectiveness on AP CSP exam performance and participation when we used an alternate set of comparison schools—one that was constant across the three analysis years and included only schools in Texas—for 12 of the Lone Star schools. In other words, the overall results were robust to the comparison schools included in analysis.

Chapter 8. Conclusion

In this chapter we list the evaluation's six research questions and present the key findings for each. We then present an overall summary and brief conclusion.

Research Question 1: Fidelity of Implementation

1. Was the Lone Star AP CSP program implemented with fidelity in
    a. its first year (the 2020–21 school year)?
    b. its second year (the 2021–22 school year)?
    c. its third year (the 2023–23 school year)?

Key findings were as follows:

- The teacher supports component was not implemented with fidelity in any of the three years.
- The student supports component was implemented with fidelity in all three years.
- The school supports component was implemented with fidelity in the first year but not in the second or third years.

Research Question 2: Participant Engagement and Perceptions

2. To what extent did Lone Star AP CSP program participants make use of the program and find it useful?

Key findings were as follows:

- Based on teachers' self-report on surveys, teacher participation in Lone Star AP CSP activities was relatively low in Year 1 and somewhat higher in Year 2. There was little change from Year 2 to Year 3 in participation levels, but perceptions of usefulness of the teacher activities increased across the three years.
- Use of the student resources, as reported by teachers, increased from year to year, as did teacher perceptions of the usefulness of these resources.
- Lack of awareness about the program's supports and resources was a commonly cited reason for nonuse of particular activities in all three years, but more so in Year 1 than Year 2, and more so in Year 2 than Year 3.
- At the close of the program in Year 3, most teachers (60%) reported that overall, they had found the program to be at least moderately useful.

- The subset of teachers who participated in all three years of the program had mixed opinions about the cohesiveness of the program across the three years, such as how well it implemented the envisioned three-year arc of professional development.

Research Question 3: Impact on Instructional Practices

3. Did participation in the Lone Star AP CSP program improve AP CSP teachers' instructional practices in the AP CSP course, compared to AP CSP teachers in the business-as-usual condition, after one year, after two years, and after three years?

Key findings were as follows:

- After one year of implementation, the Lone Star AP CSP program had a significant positive effect on the extent to which teachers encouraged students to use cognitive problem-solving strategies.
- After two years of implementation, the program had a significant and large positive effect on each of the three measures of instructional practice we examined: Engaging Students in Practices of Computer Science, Encouragement of Cognitive Problem-Solving Strategies, and Use of Pedagogically Based Learning Strategies.
- After three years of implementation, the program had no effect on any of the three measures of instructional practice.

Research Question 4: Impact on Student AP CSP Performance and Participation After One Year

4. What was the impact of one year of participation in the Lone Star AP CSP program, compared to business-as-usual AP CSP participation, on high schools' rates of (a) AP CSP exam takers receiving a qualifying score on the AP CSP exam, and (b) students taking the AP CSP exam?

Key findings were as follows:

- After one year of implementation, the Lone Star AP CSP program did not have a significant effect on schools' percentages of AP CSP exam takers that received a qualifying score on the exam.
- The program also did not have a significant effect after one year on schools' percentages of students in grades 9–12 that received a qualifying score—or a nonqualifying score—on the AP CSP exam.
- There was also no significant effect after one year of implementation on schools' percentages of students in grades 9–12 that took the AP CSP exam.

Research Question 5: Impact on Student AP CSP Performance and Participation After Two Years

5. What was the impact of *two* years of participation in the Lone Star AP CSP program, compared to business-as-usual AP CSP participation, on the same outcomes?

Key findings were as follows:

- After two years of implementation, the Lone Star AP CSP program had a borderline-significant negative effect on schools' percentages of AP CSP exam takers that received a qualifying score on the exam.
- The program had a significant negative effect on schools' percentages of students in grades 9–12 that took the AP CSP exam in the second year of implementation.
- Likely because of the negative effect on the percentages of students taking the exam, the program also had a significant negative effect on schools' percentages of students that received a qualifying score on the AP CSP exam, as well as a borderline-significant negative effect on schools' percentages of students that received a *non*qualifying score.

Research Question 6: Impact on Student AP CSP Performance and Participation After Three Years

6. What was the impact of *three* years of participation in the Lone Star AP CSP program, compared to business-as-usual AP CSP participation, on the same outcomes?

Key findings were as follows:

- After three years of implementation, the Lone Star AP CSP program did not have a significant effect on schools' percentages of AP CSP exam takers that received a qualifying score on the exam.
- The program also did not have a significant effect after three years on schools' percentages of students in grades 9–12 that received a qualifying score on the AP CSP exam, but it did have a significant effect on schools' percentages of students in grades 9–12 that received a *non*qualifying score. This effect was negative, which is in the desired direction.
- The program also had a significant negative effect on schools' percentages of students in grades 9–12 that took the AP CSP exam in the third year of implementation.

Key Takeaways

Key takeaways from the evaluation of the Lone Star AP CSP program are as follows:

- There was no year in which all three program components were implemented with fidelity, and in each of Years 2 and 3, only one of the three components was implemented with fidelity. Moreover, in none of the years was the teacher supports component—which was the program element that, as a three-year arc, made the Lone Star AP CSP program innovative and different—implemented with fidelity.
- The Lone Star AP CSP program had positive effects on participating teachers' instructional practices after the first and second years of implementation but not after the third. However, the measures of teachers' instructional practice are based on the teachers' own self-report on retrospective year-end surveys, and it is possible that teachers may have misreported or misrepresented their instructional practices (whether intentionally or not).
- In none of the three years did the program have a significant positive impact on schools' percentages of AP CSP exam takers that received a qualifying score on the exam.
- In none of the three years did the program have a significant positive impact on schools' percentages of students in grades 9–12 that took the AP CSP exam. To the contrary, in Years 2 and 3 the program's effect on student participation in the AP CSP exam was significantly negative.

Conclusion and Discussion

In comparison to matched comparison schools and controlling for baseline measures and school demographics, schools in the Lone Star AP CSP program did not display higher rates of success on or participation in the AP CSP exam in any of the three program years. Therefore, the program did not have its theorized positive effects either on AP CSP exam performance or participation. Based on survey self-report from a small number of Lone Star AP CSP teachers and matched comparison teachers, the program appears to have had a positive impact on AP CSP instructional practices in Year 2 (and to a lesser extent in Year 1), but the self-reported basis for these measures and the very small sample size, in combination with the null findings of program impact on students' AP CSP exam performance, call this instructional practice finding into question. Relatively low levels of program implementation and uptake probably help explain why the program did not achieve its hypothesized positive impacts on students. Lingering effects of the COVID-19 pandemic may have hampered implementation efforts.

It is also worth bearing in mind for the AP performance and participation findings that the "business as usual" condition for the comparison schools only means that those schools were doing whatever they would have done anyway and were not affected by the Lone Star AP CSP

program; they could have been implementing other interventions pertaining to AP CSP or computer science more generally. Numerous such interventions have been devised over the past several years amid an increasing awareness of the importance of computer science in nearly every aspect of society. Even in the EIR grant program alone, many interventions funded during the study period focused on computer science, which was considered a "competitive priority" in several of the EIR competitions.[1] Within just the cohort of 28 2019 early-phase grants awarded under Absolute Priority 3 (Science, Technology, Engineering, and Math) that the Lone Star AP CSP program belonged to, computer science was an explicit focus for 26 of them, and of those, 15 focused on or included the high school grades. At least eight (including Lone Star AP CSP) mentioned AP CSP prominently, with three (Lone Star AP CSP and two others) including it in the project title. The 2020 and 2021 EIR competitions also each had one grantee that mentioned AP CSP in the project title.[2] Given the methods used in the evaluation of the Lone Star AP CSP program and the data available, we were not able to gauge service contrast in the comparison schools, and it is quite possible that at least some of them were participating in other interventions related to AP CSP, whether funded by the EIR program or other sources such as the National Science Foundation. Against such a backdrop, significant program effects become all the more difficult to detect.

[1] The EIR early-phase and mid-phase competitions in 2019, 2020, and 2021 all had an absolute priority "with a particular focus on computer science," and the early-phase competitions from all three of those years, along with the mid-phase competition in 2021, designated computer science as a "competitive priority." See https://www.ed.gov/grants-and-programs/grants-special-populations/grants-economically-disadvantaged-students/education-innovation-and-research#eligibility-and-historical-eir-competition-applicant-info.

[2] Based on information available at https://www.ed.gov/grants-and-programs/grants-special-populations/grants-economically-disadvantaged-students/education-innovation-and-research#awards.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561–573.

Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). *Report of the 2018 NSSME+.* Horizon Research. http://horizon-research.com/NSSME/wp-content/uploads/2020/04/Report_of_the_2018_NSSME.pdf

College Board. (2018). *Guidelines for the release of data.* https://research.collegeboard.org/media/pdf/rd-guide-release-data.pdf

College Board. (n.d.-a). *AP Computer Science Principles: Research findings.* AP Central. https://apcentral.collegeboard.org/courses/ap-computer-science-principles/ap-csp-research-findings

College Board. (n.d.-b). *About AP scores.* https://apstudents.collegeboard.org/about-ap-scores

College Board. (2020). *AP Computer Science Principles course and exam description, effective fall 2020.*

Dougherty, C., Mellor, L., & Jian, S. (2006). *The relationship between Advanced Placement and college graduation.* National Center for Educational Accountability. https://files.eric.ed.gov/fulltext/ED519365.pdf

Groenwald, R. H. H., Palmer, T., & Tilling, K. (2019). *Conditioning on a mediator to adjust for unmeasured confounding.* https://osf.io/vrcuf

Holtzman, D. J. (2024). *Hand-matching tool for achieving baseline equivalence on up to four continuous variables.* American Institutes for Research.

Jones, K. T., & Holtzman, D. J. (2023, April). *Independent evaluation of Lone Star AP CSP: Implementation and uptake in the first and second years of the program.* American Institutes for Research.

Jann, B. (2017). *kmatch: Stata module for multivariate-distance and propensity-score matching, including entropy balancing, inverse probability weighting, (coarsened) exact matching, and regression adjustment.* https://ideas.repec.org/c/boc/bocode/s458346.html

Judkins, D. (2013, April). *Webinar on matching methods for i3 evaluations.* Presentation for i3 evaluators. Abt Associates.

Mattern, K. D., Marini, J. P., & Shaw, E. J. (2013). *Are AP students more likely to graduate from college on time?* The College Board. https://files.eric.ed.gov/fulltext/ED556464.pdf

Mattern, K. D., Shaw, E. J., & Ewing M. (2011). *Is AP exam participation and performance related to choice of college major?* The College Board. https://files.eric.ed.gov/fulltext/ED561044.pdf

Mattern, K. D., Shaw, E. J., & Xiong, X. (2009). *The relationship between AP exam performance and college outcomes.* The College Board. https://files.eric.ed.gov/fulltext/ED561021.pdf

Morgan, R., & Klaric, J. (2007). *AP students in college: An analysis of five-year academic careers.* The College Board.

Murphy, D., & Dodd, B. (2009). A *comparison of college performance of matched AP and non-AP student groups.* The College Board. https://files.eric.ed.gov/fulltext/ED561022.pdf

NMSI. (2019). *EIR early phase grant proposal: Lone Star AP CSP (project narrative)*. National Math + Science Initiative.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (Exp. ed.). University of Chicago Press.

VanderWeele, T. J. (2019). Principles of confounder selection. *European Journal of Epidemiology*, *34*(3), 211–219. https://doi.org/10.1007/s10654-019-00494-6

What Works Clearinghouse. (n.d). *Module 3: Baseline equivalence* [PowerPoint slides]. Institute of Education Sciences. https://ies.ed.gov/ncee/wwc/Docs/OnlineTraining/wwc_training_m3.pdf

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

What Works Clearinghouse. (2024). *Study review protocol and supplement to the* What Works Clearinghouse procedures and standards handbook, version 5.0*: Version 5.1 (December 2024).* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Document/1297#/Protocol_CurrentProtocol

Williams, T. C., Bach, C. C., Matthiesen, N. B., Henriksen, T. B., & Gagliardi, L. (2018). Directed acyclic graphs: A tool for causal studies in paediatrics. *Pediatric Research, 84,* 487–493. https://doi.org/10.1038/s41390-018-0071-3

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement.* MESA Press.

Appendix A. Supplementary Exhibits for Chapter 1

Exhibit A-1. Student Composition of the Lone Star AP CSP Schools

	Mean	Standard deviation	Minimum	Maximum
Total number of students	1,489	1,037	128	3,897
Female	46.5%	8.2%	0.0%	53.1%
Free and reduced-price lunch	68.1%	18.9%	18.7%	95.1%
Hispanic	65.7%	24.7%	18.2%	97.8%
Black or African American	15.8%	14.0%	0.6%	68.2%
White	10.3%	13.5%	1.0%	75.1%
Asian or Asian/Pacific Islander	6.7%	9.1%	0.1%	37.7%
American Indian/Alaska Native	0.2%	0.1%	0.0%	0.4%
Native Hawaiian or other Pacific Islander	0.1%	0.2%	0.0%	1.3%
Two or more races	1.2%	1.1%	0.0%	3.6%

Note. N = 37 schools for all rows.

Appendix B. Supplementary Exhibits for Chapter 2

Exhibit B-1. Fidelity Matrix and Fidelity Results Reporting Tables

a. Key Component 1 (of 3) – Support Teachers

Element (indicator)	Description	Unit of implementation	Data source(s)	Score for levels of implementation at unit level	Expected sample for fidelity measure	Expected years of fidelity measurement
Summer training	Professional development offered during summer (Summer Series preceding all three years; Content Workshop preceding Years 2 and 3)	Program	Attendance/ credit completion report provided by NMSI to AIR	Year 1: 1 = at least 80% of teachers received 16+ credits of summer training 0 = <80% of teachers received 16+ credits of summer training Year 2: 1 = at least 70% of teachers received 16+ credits of summer training 0 = <70% of teachers received 16+ credits of summer training Year 3: 1 = at least 65% of teachers attended the summer training 0 = <65% of teachers attended the summer training	Whole Lone Star AP CSP program; Year 1: 40 teachers Years 2 & 3: Teachers weighted by years of participation	2020 (Year 1) 2021 (Year 2) 2022 (Year 3)
Job-embedded courses	Synchronous virtual courses offered during the school year	Program	Attendance/ credit completion report provided by NMSI to AIR	Years 1 & 2: 1 = at least 60% of teachers received 10 credits 0 = <60% of teachers received 10 credits Year 3: 1 = at least 60% of teachers attended both the fall and spring trainings 0 = <60% of teachers attended both the fall and spring trainings	Whole Lone Star AP CSP program; Year 1: 40 teachers Years 2 & 3: Teachers weighted by years of participation	2020–21 (Y1) 2021–22 (Y2) 2022–23 (Y3)
On-demand courses	Asynchronous virtual courses	Program	Years 1, 2 & 3: Blackboard website reviewed by AIR	Years 1, 2 & 3: 1 = At least 5 courses were available 0 = <5 courses were available	Whole Lone Star AP CSP program	2020–21 (Y1) 2021–22 (Y2) 2022–23 (Y3)
Key component score				Y1: Sum of program-level indicator scores: Possible range is 0–3 Adequate = 2+	**Y1 score: 1**	2020–21
				Y2: Sum of program-level indicator scores: Possible range is 0–3 Adequate = 2+	**Y2 score: 1**	2021–22
				Y3: Sum of program-level indicator scores: Possible range is 0–-3 Adequate = 2+	**Y3 score: 1**	2022–23
FIDELITY DETERMINATION				**Met threshold for implementation with fidelity?**	**Y1:** ○ Yes ◉ No	
					Y2: ○ Yes ◉ No	
					Y3: ○ Yes ◉ No	

b. Key Component 2 (of 3) – Support Students

Element (indicator)	Description	Unit of implementation	Data source(s)	Score for levels of implementation at unit level	Expected sample for fidelity measure	Expected years of fidelity measurement
Synchronous supports	Live virtual study sessions split between 12 topics; optional for students	Program	Schedule of synchronous student supports that occurred, including which topic was addressed and how many students attended; provided by NMSI to AIR	Years 1, 2, & 3: 1 = at least one study session per topic was available to all students 0 = at least one study session per topic was not available to students	Whole Lone Star AP CSP program	2020–21 (Y1) 2021–22 (Y2) 2022–23 (Y3)
On-demand (asynchronous) supports	Online library of modules; video recordings from NMSI coaches with office hours available; optional for students	Program	Blackboard website; schedule of office hour sessions that occurred, including which topic was addressed and how many students attended; reviewed by AIR	Years 1, 2, & 3: 1 = at least one online module and one office hour session per topic was available to students 0 = at least one online module and one office hour session per topic was not available	Whole Lone Star AP CSP program	2020–21 (Y1) 2021–22 (Y2) 2022–23 (Y3)
Exam fee subsidies	Students are given a subsidy for the AP exam fee (value dependent on the student's free and reduced-price lunch status)	Program	Subsidy payment vouchers/distribution report provided by NMSI to AIR; if any schools decline the subsidy, documentation of this	Years 1, 2, & 3: 1 = all schools that requested subsidies received their requested amount 0 = not all schools that requested subsidies received subsidies	Whole Lone Star AP CSP program, 37 schools	2020–21 (Y1) 2021–22 (Y2) 2022-23 (Y3)
Key Component Score				Y1: Sum of program-level indicator scores: Possible range is 0–3 Adequate = 3	**Y1 score: 3**	2020–21
				Y2: Sum of program-level indicator scores: Possible range is 0–3 Adequate = 3	**Y2 score: 3**	2021–22
				Y3: Sum of program-level indicator scores: Possible range is 0–3 Adequate = 3	**Y3 score: 3**	2022–23
FIDELITY DETERMINATION				**Met threshold for implementation with fidelity?**	**Y1:** ◉ Yes ○ No	
					Y2: ◉ Yes ○ No	
					Y3: ◉ Yes ○ No	

c. Key Component 3 (of 3) – Support Schools

Element (indicator)	Description	Unit of implementation	Data source(s)	Score for levels of implementation at unit level	Expected sample for fidelity measure	Expected years of fidelity measurement
Technology, equipment, and supplies	Schools receive laptops and/or tablets	Program	Shipping receipts/records; emails from schools that declined supplies; provided by NMSI to AIR	Year 1: 1 = 100% of schools that requested equipment received it from NMSI 0 = <100% of schools that requested equipment received it from NMSI Years 2 and 3: not applicable	Whole Lone Star AP CSP program; 37 schools	2020 (Y1)
Leadership training	Ongoing training for school leadership	Program	Attendance report(s) provided by NMSI to AIR	Year 3 (cumulative for all 3 years): 1 = at least 75% of schools had at least one leader participate in leadership training for at least one summer 0 = <75% of schools had at least one leader participate in leadership training for at least one summer	Whole Lone Star AP CSP program; 37 schools	2022–23 (Y3)
Technical assistance	All programmatic support from NMSI, including regular meetings with the school director, site coordinator, and a counselor	Program	Documentation of outreach to schools (e.g., copy of email, call log) provided by NMSI to AIR	Years 1, 2, & 3: 1 = at least 80% of schools met with NMSI at least five times, including once in fall (Aug–Nov), once in winter (Dec–Feb), and once in spring (Mar–May) 0 = <80% of schools met with NMSI as described	Whole Lone Star AP CSP program; 37 schools	2020–21 (Y1) 2021–22 (Y2) 2022–23 (Y3)
Counselor training/ online coaching	Coaching sessions conducted virtually but synchronously (seven 1-hour sessions in Year 2 and three multi-hour sessions in Year 3)	Program	Attendance records provided by NMSI to AIR	Year 1: Not part of the program Years 2 & 3: 1 = at least 80% of schools had at least one counselor attend 5+ hours of coaching 0 = <80% of schools had at least one counselor attend 5+ hours of coaching	Whole Lone Star AP CSP program; 37 schools	2021–22 (Y2) 2022–23 (Y3)
Stipends for school staff	Monetary incentive for school director ($500), site coordinator ($500), and counselor ($1,000)	Program	Documentation of payment (e.g., receipts, email confirmation) provided by NMSI to AIR	Years 1, 2, & 3: 1 = at least 80% of schools received stipends totaling $2,000 0 = <80% of schools received stipends totaling $2,000	Whole Lone Star AP CSP program; 37 schools	2020–21 (Y1) 2021–22 (Y2) 2022–23 (Y3)
Key Component Score				Y1: Sum of program-level indicator scores: Possible range is 0–3 Adequate = 2+	**Y1 score: 2**	2020–21
				Y2: Sum of program-level indicator scores: Possible range is 0–3 Adequate = 2+	**Y2 score: 1**	2021–22
				Y3: Sum of program-level indicator scores: Possible range is 0–4 Adequate = 3+	**Y3 score: 1**	2022–23
FIDELITY DETERMINATION				**Met threshold for implementation with fidelity?**	**Y1:** ◉ Yes ○ No	
					Y2: ○ Yes ◉ No	
					Y3: ○ Yes ◉ No	

Exhibit B-2. Calculation of Weighted Percentages for Year 2 Teacher Supports Component Elements 1 and 2

Activity	Teachers continuing from Year 1 to Year 2		Teachers starting in Year 2		Weighted percentage (weighted total/78)
	Unweighted	Weighted (unweighted × 2)	Unweighted	Weighted (unweighted × 1)	
In the sample	34	68	10	10	100.00%
Attended any summer training (Content Workshop and/or Summer Series)	14	28	3	3	39.74%
Met 16-credit threshold for summer training	11	22	0	0	28.21%
Registered for any job-embedded training	17	34	3	3	47.44%
Met 10-credit threshold for job-embedded training	6	12	1	1	16.67%

Exhibit B-3. Calculation of Weighted Percentages for Year 3 Teacher Supports Component Elements 1 and 2

Activity	Teachers continuing from Year 1 through Year 3		Teachers continuing from Year 2 to Year 3		Teachers starting in Year 3		Weighted percentage (weighted total/112)
	Unweighted	Weighted (unweighted × 3)	Unweighted	Weighted (unweighted × 2)	Unweighted	Weighted (unweighted × 1)	
In the sample	27	81	10	20	11	11	100.00%
Attended summer training (i.e., met threshold)	11	33	6	12	6	6	45.54%
Attended job-embedded training in *either* fall or spring	12	36	5	10	9	9	49.11%
Attended job-embedded training in *both* fall and spring (i.e., met threshold)	10	30	5	10	8	8	42.86%

Appendix C. Supplementary Exhibits for Chapter 4

Exhibit C-1. Correlations Among the Three Instructional Practice Measures

	Engaging Students in Practices of Computer Science	Encouragement of Cognitive Problem-Solving Strategies	Use of Pedagogically Based Learning Strategies
Engaging Students in Practices of Computer Science	1.000		
Encouragement of Cognitive Problem-Solving Strategies	0.677	1.000	
Use of Pedagogically Based Learning Strategies	0.734	0.673	1.000

Note. These correlations were calculated using the 2020 scaling sample. *N* = 102 for the first measure and 103 for the second and third measures.

Exhibit C-2. Teacher Characteristics for the Year 1 Analysis Sample, Continuous Variables

Variable	Lone Star teachers (*n* = 9)	Comparison teachers (*n* = 54)	*p* value	Hedges' *g*
Total years of teaching experience (any subject, any grade level) as of the 2019–20 school year	9.000 (6.801)	15.963 (7.787)	0.014	0.897
Years of experience teaching computer science as of the 2019–20 school year	3.667 (4.555)	6.731 (6.612)	0.187	0.474
Number of years the teacher had held a full-time job in computer science not related to teaching or education[a]	0.111 (0.220)	3.222 (6.773)	0.176	0.487
Teacher's age (as of spring 2021)	38.000 (8.846)	47.759 (8.988)	0.004	1.075

[a] This variable was coded as 0 for teachers who had never held such a job.

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from *t* tests.

Exhibit C-3. Teacher Characteristics for the Year 1 Analysis Sample, Dichotomous Variables

Variable	Lone Star teachers (*n* = 9)	Comparison teachers (*n* = 54)	*p* value	Cox effect size
Had more than 2 years of CS teaching experience as of the 2019–20 school year	44.44%	81.48%	0.024	-1.020
Had more than 3 years of CS teaching experience as of the 2019–20 school year	22.22%	62.96%	0.036	-1.067
Holds master's degree in a STEM field	44.44%	18.52%	0.096	0.753
Certified to teach computer science	55.56%	68.52%	0.449	-0.332
Certified to teach mathematics	66.67%	53.70%	0.472	0.326
Ever held a full-time job in computer science not related to teaching or education	22.22%	29.63%	0.650	-0.232
Female	22.22%	70.37%	0.013	-1.268

Note. The *p* values in the third column are from logistic regression. See What Works Clearinghouse (n.d.) for the formula for the Cox effect size.

Exhibit C-4. AP CSP Student Characteristics in the Classrooms of the Teachers in the Year 1 Analysis Sample, Continuous Variables

Variable	Lone Star teachers	Comparison teachers	*p* value	Hedges' *g*
Number of students/sections taught				
Number of students enrolled in teacher's AP CSP class(es) in fall 2019	28.125 (18.527) 8	41.315 (38.350) 54	0.345	0.356
Number of students enrolled in teacher's AP CSP class(es) in spring 2020	25.375 (14.382) 8	37.942 (37.087) 52	0.350	0.353
Number of AP CSP sections the teacher taught in 2019–20	1.333 (0.500) 9	1.685 (0.968) 54	0.292	0.378
Proportion of 2019–20 AP CSP students in each grade level				
9th grade	0.393 (0.406) 8	0.098 (0.212) 54	0.002	-1.196
10th grade	0.241 (0.181) 8	0.242 (0.220) 54	0.991	0.004

Variable	Lone Star teachers	Comparison teachers	*p* value	Hedges' *g*
11th grade	0.259 (0.284) 8	0.288 (0.195) 54	0.709	0.140
12th grade	0.108 (0.119) 8	0.353 (0.249) 54	0.008	1.021
Other grade level	0.000 (0.000) 8	0.019 (0.136) 54	0.701	0.144
Proportion of 2019–20 AP CSP students in each gender category				
Female	0.226 (0.132) 7	0.328 (0.217) 53	0.229	0.482
Male	0.774 (0.132) 7	0.669 (0.218) 53	0.220	-0.492
Non-binary/other/not sure	0.000 (0.000) 7	0.002 (0.009) 53	0.517	0.259
Proportion of 2019–20 AP CSP students in race/ethnicity categories				
Asian	0.112 (0.087) 6	0.081 (0.155) 53	0.633	-0.204
Black or African American	0.080 (0.098) 6	0.105 (0.166) 53	0.717	0.155
Hispanic or Latinx	0.600 (0.318) 6	0.190 (0.260) 53	0.001	-1.523
White	0.100 (0.089) 6	0.492 (0.290) 53	0.002	1.389
All other ethnicities or unknown ethnicity[a]	0.108 (0.132) 6	0.132 (0.202) 53	0.781	0.119

[a] This category includes Middle Eastern or North African, Native Hawaiian or other Pacific Islander, multiracial/multiethnic, other race/ethnicity, and "not sure."

Note. In the first two columns, each cell contains the mean, the standard deviation (in parentheses), and the number of teachers reporting. The *p* values in the third column are from *t* tests.

Exhibit C-5. AP CSP Student Characteristics in the Classrooms of the Teachers in the Year 1 Analysis Sample, Dichotomous Variables

Variable	Lone Star teachers	Comparison teachers	*p* value	Cox effect size
Less than or equal to 5% of 2019–20 AP CSP students are English language learners	16.67% 6	64.15% 53	0.053	-1.311
More than 10% of 2019–20 AP CSP students are English language learners	66.67% 6	13.21% 53	0.007	1.541
More than 50% of 2019–20 AP CSP students are Hispanic/Latinx	50.00% 6	11.32% 53	0.026	1.231
More than 50% of 2019–20 AP CSP students are Hispanic/Latinx, with imputation for missing values using 2020–21 data	55.56% 9	11.11% 54	0.004	1.378

Note. In the first two columns, each cell contains the percentage and the number of teachers included in the calculation of the percentage. The *p* values in the third column are from logistic regression. See What Works Clearinghouse (n.d.) for the formula for the Cox effect size.

Exhibit C-6. Teacher Characteristics for the Year 2 Analysis Sample, Continuous Variables

Variable	Lone Star teachers	Comparison teachers	*p* value	Hedges' g
Total years of teaching experience (any subject, any grade level) as of the 2019–20 school year	12.000 (8.226) 7	16.837 (7.512) 49	0.121	0.628
Years of experience teaching computer science as of the 2019–20 school year	5.286 (7.135) 7	6.898 (6.329) 49	0.537	0.247
Number of years the teacher had held a full-time job in computer science not related to teaching or education[a]	0.357 (0.748) 7	3.000 (6.916) 49	0.321	0.399
Teacher's age (as of spring 2021)	37.200 (11.411) 5	48.587 (7.932) 46	0.005	1.356

[a] This variable was coded as 0 for teachers who had never held such a job.

Note. In the first two columns, each cell contains the mean, the standard deviation (in parentheses), and the number of teachers reporting. The *p* values in the third column are from *t* tests.

Exhibit C-7. Teacher Characteristics for the Year 2 Analysis Sample, Dichotomous Variables

Variable	Lone Star teachers (n = 7)	Comparison teachers (n = 49)	*p* value	Cox effect size
Had more than 2 years of CS teaching experience as of the 2019–20 school year	57.14%	85.71%	0.082	-0.899
Had more than 3 years of CS teaching experience as of the 2019–20 school year	28.57%	67.35%	0.065	-0.980
Holds master's degree in a STEM field	28.57%	22.45%	0.721	0.193
Certified to teach computer science	71.43%	65.31%	0.750	0.170
Certified to teach mathematics	57.14%	57.14%	1.000	0.000
Ever held a full-time job in computer science not related to teaching or education	28.57%	20.41%	0.625	0.266
Female	28.57%	67.35%	0.065	-0.980

Note. The *p* values in the third column are from logistic regression. See What Works Clearinghouse (n.d.) for the formula for the Cox effect size.

Exhibit C-8. AP CSP Student Characteristics in the Classrooms of the Teachers in the Year 2 Analysis Sample, Continuous Variables

Variable	Lone Star teachers	Comparison teachers	*p* value	Hedges' g
Number of students/sections taught				
Number of students enrolled in teacher's AP CSP class(es) in fall 2019	33.429 (19.646) 7	39.939 (39.067) 49	0.668	0.172
Number of students enrolled in teacher's AP CSP class(es) in spring 2020	30.000 (16.482) 7	38.340 (39.203) 47	0.583	0.220
Number of AP CSP sections the teacher taught in 2019–20	1.571 (0.535) 7	1.714 (0.957) 49	0.702	0.153
Proportion of 2019–20 AP CSP students in each grade level				
9th grade	0.161 (0.216) 7	0.107 (0.220) 49	0.543	-0.244
10th grade	0.337 (0.145) 7	0.253 (0.204) 49	0.300	-0.417

Variable	Lone Star teachers	Comparison teachers	*p* value	Hedges' g
11th grade	0.326 (0.259) 7	0.269 (0.177) 49	0.456	-0.299
12th grade	0.176 (0.165) 7	0.371 (0.257) 49	0.058	0.773
Proportion of 2019–20 AP CSP students in each gender category				
Female	0.237 (0.139) 6	0.332 (0.215) 48	0.296	0.450
Male	0.763 (0.139) 6	0.666 (0.215) 48	0.289	-0.457
Non-binary/other/not sure	0.000 (0.000) 6	0.001 (0.007) 48	0.589	0.232
Proportion of 2019–20 AP CSP students in race/ethnicity categories				
Asian	0.220 (0.237) 5	0.093 (0.174) 48	0.140	-0.694
Black or African American	0.044 (0.046) 5	0.105 (0.186) 48	0.474	0.334
Hispanic or Latinx	0.586 (0.322) 5	0.165 (0.245) 48	0.001	-1.645
White	0.048 (0.087) 5	0.509 (0.308) 48	0.002	1.532
All other ethnicities or unknown ethnicity[a]	0.102 (0.116) 5	0.129 (0.212) 48	0.785	0.127

[a] This category includes Middle Eastern or North African, Native Hawaiian or other Pacific Islander, multiracial/multiethnic, other race/ethnicity, and "not sure."

Note. In the first two columns, each cell contains the mean, the standard deviation (in parentheses), and the number of teachers reporting. The *p* values in the third column are from *t* tests.

Exhibit C-9. AP CSP Student Characteristics in the Classrooms of the Teachers in the Year 2 Analysis Sample, Dichotomous Variables

Variable	Lone Star teachers	Comparison teachers	*p* value	Cox effect size
Less than or equal to 5% of 2019–20 AP CSP students are English language learners	20.00% 5	62.50% 48	0.101	-1.133
More than 10% of 2019–20 AP CSP students are English language learners	60.00% 5	16.67% 48	0.042	1.203
More than 50% of 2019–20 AP CSP students are Hispanic/Latinx	40.00% 5	8.33% 48	0.058	1.190
More than 50% of 2019–20 AP CSP students are Hispanic/Latinx, with imputation for missing values using 2020–21 data	40.00% 5	8.16% 49	0.055	1.203

Note. In the first two columns, each cell contains the percentage and the number of teachers included in the calculation of the percentage. The *p* values in the third column are from logistic regression. See What Works Clearinghouse (n.d.) for the formula for the Cox effect size.

Exhibit C-10. Teacher Characteristics for the Year 3 Analysis Sample, Continuous Variables

Variable	Lone Star teachers (*n* = 7)	Comparison teachers (*n* = 35)	*p* value	Hedges' *g*
Total years of teaching experience (any subject, any grade level) as of the 2019–20 school year	13.714 (7.228)	17.229 (7.934)	0.285	0.440
Years of experience teaching computer science as of the 2019–20 school year	5.000 (7.280)	7.514 (6.568)	0.369	0.369
Number of years the teacher had held a full-time job in computer science not related to teaching or education[a]	0.286 (0.756)	3.714 (7.061)	0.211	0.516
Teacher's age (as of spring 2023)	47.857 (13.874)	51.543 (8.438)	0.352	0.382

[a] This variable was coded as 0 for teachers who had never held such a job.

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from *t* tests.

Exhibit C-11. Teacher Characteristics for the Years 3 Analysis Sample, Dichotomous Variables

Variable	Lone Star teachers (*n* = 7)	Comparison teachers (*n* = 35)	*p* value	Cox effect size
Had more than 2 years of CS teaching experience as of the 2019–20 school year	42.86%	85.71%	0.021	-1.236
Had more than 3 years of CS teaching experience as of the 2019–20 school year	28.57%	77.14%	0.022	-1.268
Holds master's degree in a STEM field	14.29%	22.86%	0.618	-0.342
Certified to teach computer science	71.43%	71.43%	1.000	0.000
Certified to teach mathematics	71.43%	54.29%	0.410	0.443
Ever held a full-time job in computer science not related to teaching or education	14.29%	28.57%	0.444	-0.521
Female	14.29%	65.71%	0.032	-1.452

Note. The *p* values in the third column are from logistic regression. See What Works Clearinghouse (n.d.) for the formula for the Cox effect size.

Exhibit C-12. AP CSP Student Characteristics in the Classrooms of the Teachers in the Year 3 Analysis Sample, Continuous Variables

Variable	Lone Star teachers	Comparison teachers	*p* value	Hedges' *g*
Number of students/sections taught				
Number of students enrolled in teacher's AP CSP class(es) in fall 2019	33.857 (19.126) 7	41.171 (41.981) 35	0.656	0.182
Number of students enrolled in teacher's AP CSP class(es) in spring 2020	30.857 (15.323) 7	39.515 (42.510) 33	0.601	0.215
Number of AP CSP sections the teacher taught in 2019–20	1.571 (0.535) 7	1.686 (0.932) 35	0.756	0.127
Proportion of 2019–20 AP CSP students in each grade level				
9th grade	0.304 (0.369) 7	0.095 (0.198) 35	0.035	-0.885
10th grade	0.323 (0.174) 7	0.279 (0.186) 35	0.572	-0.232

Variable	Lone Star teachers	Comparison teachers	*p* value	Hedges' *g*
11th grade	0.211 (0.179) 7	0.305 (0.173) 35	0.200	0.530
12th grade	0.161 (0.176) 7	0.320 (0.183) 35	0.042	0.855
Proportion of 2019–20 AP CSP students in each gender category				
Female	0.170 (0.063) 6	0.297 (0.156) 34	0.058	0.848
Male	0.830 (0.063) 6	0.701 (0.156) 34	0.055	-0.859
Non-binary/other/not sure	0.000 (0.000) 6	0.002 (0.008) 34	0.521	0.281
Proportion of 2019–20 AP CSP students in race/ethnicity categories				
Asian	0.220 (0.237) 5	0.109 (0.186) 34	0.235	-0.566
Black or African American	0.044 (0.046) 5	0.076 (0.130) 34	0.591	0.254
Hispanic or Latinx	0.626 (0.321) 5	0.131 (0.148) 34	0.000	-2.766
White	0.028 (0.044) 5	0.544 (0.292) 34	0.000	1.831
All other ethnicities or unknown ethnicity[a]	0.082 (0.125) 5	0.140 (0.248) 34	0.615	0.238

[a] This category includes Middle Eastern or North African, Native Hawaiian or other Pacific Islander, multiracial/multiethnic, other race/ethnicity, and "not sure."

Note. In the first two columns, each cell contains the mean, the standard deviation (in parentheses), and the number of teachers reporting. The *p* values in the third column are from *t* tests.

Exhibit C-13. AP CSP Student Characteristics in the Classrooms of the Teachers in the Year 3 Analysis Sample, Dichotomous Variables

Variable	Lone Star teachers	Comparison teachers	*p* value	Cox effect size
Less than or equal to 5% of 2019–20 AP CSP students are English language learners	0.00% 5	67.65% 34	n/a	n/a
More than 10% of 2019–20 AP CSP students are English language learners	80.00% 5	5.88% 34	0.002	2.469
More than 50% of 2019–20 AP CSP students are Hispanic/Latinx	60.00% 5	0.00% 34	n/a	n/a
More than 50% of 2019–20 AP CSP students are Hispanic/Latinx, with imputation for missing values using 2020–21 data	71.43% 7	0.00% 35	n/a	n/a

Note. In the first two columns, each cell contains the percentage and the number of teachers included in the calculation of the percentage. The *p* values in the third column are from logistic regression. See What Works Clearinghouse (n.d.) for the formula for the Cox effect size. Logistic regression and calculation of the Cox effect size are only possible when both groups have a non-zero percentage.

Exhibit C-14. Full Model Regression Results for Analyses of Teachers' Instructional Practices After One Year of Implementation

Variable	Engaging Students in Practices of Computer Science	Encouragement of Cognitive Problem-Solving Strategies	Use of Pedagogically Based Learning Strategies
Lone Star program participation	0.6650 (0.4053) 0.106	1.2913* (0.5819) 0.030	0.4897 (0.3904) 0.215
Baseline measure of the outcome	0.6149*** (0.1202) 0.000	0.4795*** (0.1244) 0.000	0.6481*** (0.1269) 0.000
Number of years held full-time job in CS not related to teaching or education	-0.0158 (0.0216) 0.468	-0.0242 (0.0312) 0.442	-0.0257 (0.0208) 0.222
Has master's degree in a STEM field	0.4735 (0.3269) 0.153	0.6102 (0.4630) 0.193	0.8222* (0.3098) 0.010
Is female	0.2937 (0.2896) 0.315	0.8035† (0.4143) 0.057	0.4788† (0.2848) 0.098
Constant	-0.4827† (0.2658) 0.075	-0.4245 (0.3926) 0.284	-1.0867*** (0.2552) 0.000
Observations	63	63	63
R-squared	0.398	0.320	0.439

***$p < .001$, **$p < .01$, *$p < .05$, †$p < .10$

Note. Each cell contains the coefficient, the standard error (in parentheses), and the *p* value.

Exhibit C-15. Full Model Regression Results for Analyses of Teachers' Instructional Practices After Two Years of Implementation

Variable	Engaging Students in Practices of Computer Science	Encouragement of Cognitive Problem-Solving Strategies	Use of Pedagogically Based Learning Strategies
Lone Star program participation	2.1064*** (0.4310) 0.000	1.6493** (0.4888) 0.001	1.6162*** (0.3368) 0.000
Baseline measure of the outcome	0.8337*** (0.1104) 0.000	0.6951*** (0.1084) 0.000	0.8215*** (0.1019) 0.000
Number of years held full-time job in CS not related to teaching or education	0.0038 (0.0228) 0.869	-0.0235 (0.0257) 0.364	0.0066 (0.0175) 0.707
Constant	0.2618 (0.1651) 0.119	0.7176** (0.2105) 0.001	-0.0133 (0.1285) 0.918
Observations	56	56	56
R-squared	0.627	0.530	0.624

***$p < .001$, **$p < .01$, *$p < .05$, †$p < .10$

Note. Each cell contains the coefficient, the standard error (in parentheses), and the *p* value.

Exhibit C-16. Full Model Regression Results for Analyses of Teachers' Instructional Practices After Three Years of Implementation

Variable	Engaging Students in Practices of Computer Science	Encouragement of Cognitive Problem-Solving Strategies	Use of Pedagogically Based Learning Strategies
Lone Star program participation	0.0939 (0.3792) 0.806	0.0344 (0.3878) 0.930	0.2449 (0.2686) 0.368
Baseline measure of the outcome	0.5772*** (0.1062) 0.000	0.6761*** (0.1009) 0.000	0.6956*** (0.0899) 0.000
Number of years teaching computer science	-0.0109 (0.0224) 0.629	-0.0545* (0.0219) 0.018	-0.0090 (0.0150) 0.553
Number of years held full-time job in CS not related to teaching or education	-0.0234 (0.0228) 0.313	-0.0492* (0.0226) 0.036	-0.0308† (0.0157) 0.058
Constant	0.1491 (0.2433) 0.544	0.8451** (0.2414) 0.001	-0.1235 (0.1663) 0.462
Observations	42	42	42
R-squared	0.527	0.620	0.677

***$p < .001$, **$p < .01$, *$p < .05$, †$p < .10$

Note. Each cell contains the coefficient), the standard error (in parentheses), and the *p* value.

Appendix D. Supplementary Exhibits for Chapter 5

Exhibit D-1. Within-Block Statistics on 2019–20 Enrollment in Grades 9–12, All Blocks in the Year 1 Analysis Sample

Subsample/block	Lone Star schools	Comparison schools	*p* value	Hedges' *g*
Subsample 1	1,526 (1,207) 13	1,600 (1,097) 130	0.819	-0.066
Subsample 2, Block 2a	1,306 (1,252) 11	1,353 (871) 33	0.892	-0.047
Subsample 2, Block 2b	988 (821) 4	966 (685) 24	0.954	0.031

Note. In the first two columns, each cell contains the mean, the standard deviation (in parentheses), and the number of schools. The *p* values in the third column are from *t* tests.

Exhibit D-2. Within-Block Baseline Statistics for the Year 1, Subsample 2 Analysis Sample

Measure	Lone Star schools	Comparison schools	*p* value	Hedges' *g*
Block 2a (11 Lone Star schools and 33 comparison schools)				
Average percentage (across all 38 AP subject area exams) of 2020 exam takers earning a 3, 4, or 5	37.04% (15.10)	38.28% (15.94)	0.822	-0.077
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 3, 4, or 5 in 2020	0.32% (0.25)	0.36% (0.32)	0.678	-0.143
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 1 or 2 in 2020	0.54% (0.43)	0.55% (0.39)	0.957	-0.019
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 taking a 2020 AP exam	0.86% (0.55)	0.91% (0.61)	0.803	-0.086
Block 2b (4 Lone Star schools and 24 comparison schools)				
Average percentage (across all 38 AP subject area exams) of 2020 exam takers earning a 3, 4, or 5	35.28% (19.64)	35.48% (17.50)	0.983	-0.011
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 3, 4, or 5 in 2020	0.28% (0.35)	0.28% (0.30)	0.970	0.020
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 1 or 2 in 2020	0.34% (0.23)	0.34% (0.20)	0.989	0.007
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 taking a 2020 AP exam	0.63% (0.57)	0.62% (0.49)	0.977	0.015

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from *t* tests.

Exhibit D-3. 2020–21 School Characteristics for the Year 1, Subsample 1 Analysis Sample, Continuous Variables

Variable	Lone Star schools (*n* = 13)	Comparison schools (*n* = 130)	*p* value	Hedges' *g*
Total number of students enrolled	1,614 (1,096)	1,679 (1,059)	0.832	-0.061
Percentage of students who are Black	12.1% (10.1)	12.2% (14.8)	0.972	-0.010
Percentage of students who are Hispanic	65.0% (24.7)	33.5% (25.4)	0.000	1.238
Percentage of students who are female	47.4% (3.2)	49.4% (6.0)	0.253	-0.332
Within-state standardized percentage of students who are economically disadvantaged (missing values = 0)	0.21 (0.70)	-0.19 (1.00)	0.157	0.412

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from *t* tests.

Exhibit D-4. 2020–21 School Characteristics for the Year 1, Subsample 1 Analysis Sample, Dichotomous Variables

Variable	Lone Star schools (*n* = 13)	Comparison schools (*n* = 130)	*p* value	Cox effect size
Charter school	38.5%	6.9%	0.001	1.283
Locale type city	53.8%	23.8%	0.027	0.793
Locale type suburb	30.8%	56.9%	0.082	-0.657
Locale type rural or town	15.4%	19.2%	0.736	-0.163
Missing percentage of students who are economically disadvantaged	0.0%	0.8%	n/a	n/a

Note. The *p* values in the third column are from logistic regression. See What Works Clearinghouse (n.d.) for the formula for the Cox effect size.

Exhibit D-5. 2020–21 School Characteristics for the Year 1, Subsample 2 Analysis Sample, Continuous Variables

Variable	Lone Star schools (*n* = 15)	Comparison schools (*n* = 57)	*p* value	Hedges' *g*
Total number of students enrolled	1,409 (1,000)	1,267 (787)	0.559	0.169
Percentage of students who are Black	20.3% (16.7)	13.7% (18.8)	0.219	0.356
Percentage of students who are Hispanic	58.2% (27.5)	53.2% (33.2)	0.591	0.155
Percentage of students who are female	44.6% (12.4)	49.5% (5.7)	0.027	-0.65
Within-state standardized percentage of students who are economically disadvantaged	0.16 (0.81)	0.35 (0.97)	0.481	-0.203

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from *t* tests.

Exhibit D-6. 2020–21 School Characteristics for the Year 1, Subsample 2 Analysis Sample, Dichotomous Variables

Variable	Lone Star schools (*n* = 15)	Comparison schools (*n* = 57)	*p* value	Cox effect size
Charter school	60.0%	15.8%	0.001	1.247
Locale type city	60.0%	47.4%	0.387	0.306
Locale type suburb	40.0%	19.3%	0.101	0.615
Locale type rural or town	0.0%	33.3%	n/a	n/a
Missing percentage of students who are economically disadvantaged	0.0%	0.0%	n/a	n/a

Note. The *p* values in the third column are from logistic regression. See What Works Clearinghouse (n.d.) for the formula for the Cox effect size.

Exhibit D-7. Full Model Regression Results for the Year 1, Subsample 1 Analyses

Variable	Proportion of 2021 AP CSP exam takers earning a 3, 4, or 5 on the exam	Proportion of students in grades 9–12 earning a 3, 4, or 5 on the 2021 AP CSP exam	Proportion of students in grades 9–12 earning a 1 or 2 on the 2021 AP CSP exam	Proportion of students in grades 9–12 taking the 2021 AP CSP exam
Lone Star school	-0.0011 (0.0617) 0.986	0.0015 (0.0031) 0.632	0.0028 (0.0036) 0.435	0.0041 (0.0056) 0.463
Baseline measure of the outcome	0.5284*** (0.0930) 0.000	0.5660*** (0.1022) 0.000	0.9947*** (0.0947) 0.000	0.8971*** (0.1003) 0.000
Charter school	-0.0675 (0.0620) 0.278	-0.0018 (0.0033) 0.594	0.0100* (0.0039) 0.012	0.0068 (0.0061) 0.273
Total number of students enrolled	0.0000 (0.0000) 0.306	-0.0000** (0.0000) 0.005	-0.0000 (0.0000) 0.397	-0.0000 (0.0000) 0.119
Locale type suburb	0.0294 (0.0413) 0.478	-0.0028 (0.0021) 0.173	-0.0031 (0.0024) 0.199	-0.0049 (0.0036) 0.177
Locale type rural or town	0.0749 (0.0531) 0.161	0.0016 (0.0026) 0.548	-0.0048 (0.0030) 0.120	-0.0023 (0.0047) 0.618
Proportion of students who are female	0.0617 (0.2814) 0.827	-0.0189 (0.0145) 0.195	-0.0122 (0.0162) 0.452	-0.0361 (0.0253) 0.156
Proportion of students who are Black	-0.2145 (0.1347) 0.114	-0.0012 (0.0067) 0.857	0.0048 (0.0075) 0.526	0.0061 (0.0117) 0.606
Proportion of students who are Hispanic	-0.2343** (0.0895) 0.010	-0.0041 (0.0043) 0.339	-0.0000 (0.0051) 0.997	-0.0010 (0.0076) 0.897
Within-state standardized proportion of students who are economically disadvantaged	-0.0400† (0.0217) 0.068	-0.0020† (0.0011) 0.071	0.0002 (0.0012) 0.871	-0.0016 (0.0019) 0.412
Missing proportion of students who are economically disadvantaged	-0.0343 (0.1910) 0.858	-0.0094 (0.0098) 0.340	0.0007 (0.0109) 0.948	-0.0129 (0.0170) 0.448
Constant	0.3114† (0.1699) 0.069	0.0228** (0.0079) 0.004	0.0102 (0.0088) 0.246	0.0286* (0.0137) 0.038
Observations	143	143	143	143
R-squared	0.514	0.482	0.673	0.608

***p < .001, **p < .01, *p < .05, †p < .10

Note. Each cell contains the coefficient, the standard error (in parentheses), and the p value.

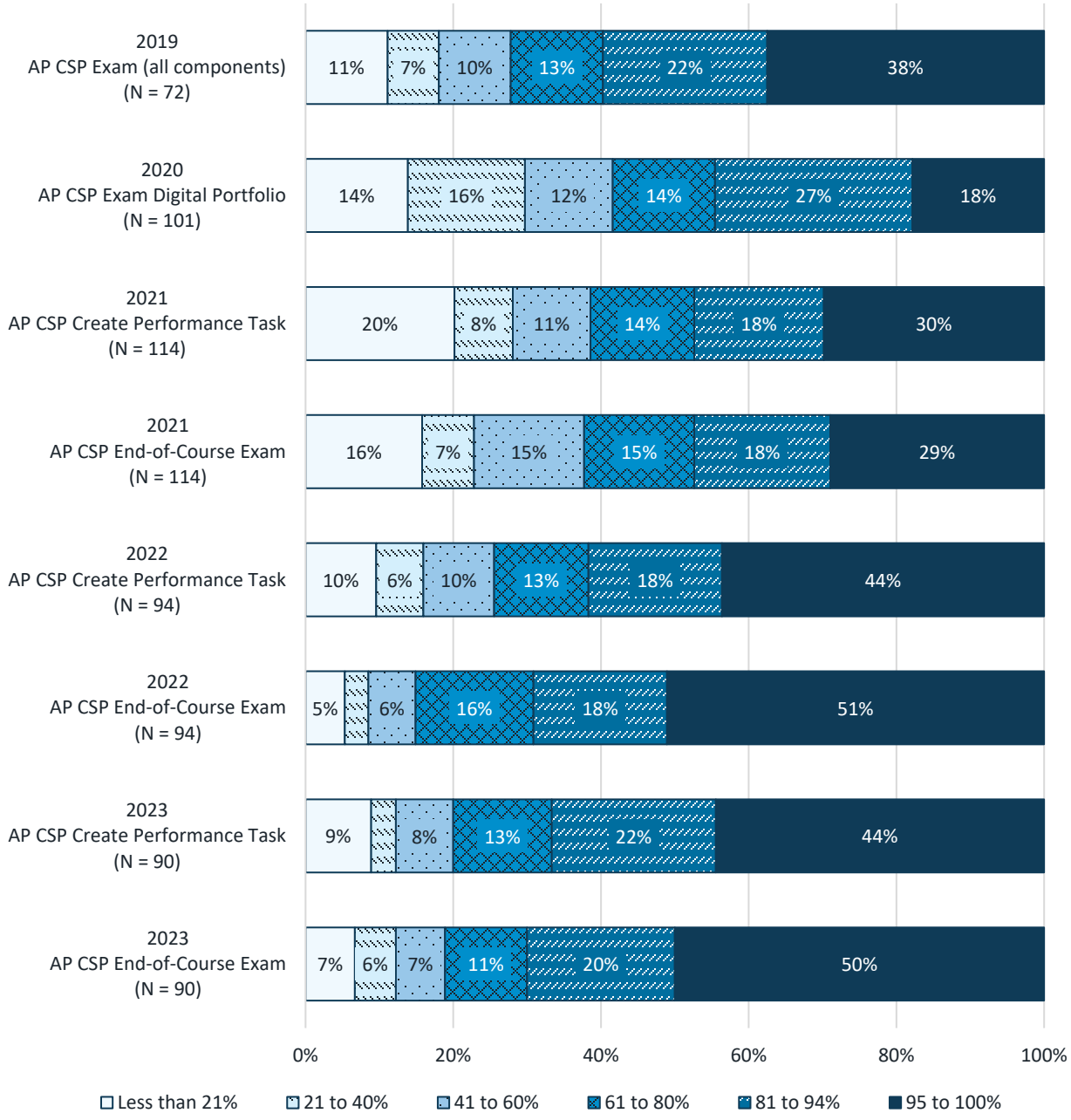
Exhibit D-8. Full Model Regression Results for the Year 1, Subsample 2 Analyses

Variable	Proportion of 2021 AP CSP exam takers earning a 3, 4, or 5 on the exam	Proportion of students in grades 9–12 earning a 3, 4, or 5 on the 2021 AP CSP exam	Proportion of students in grades 9–12 earning a 1 or 2 on the 2021 AP CSP exam	Proportion of students in grades 9–12 taking the 2021 AP CSP exam
Lone Star school	-0.0551 (0.0601) 0.363	0.0024 (0.0046) 0.605	0.0071 (0.0116) 0.543	0.0022 (0.0152) 0.887
Baseline measure of the outcome	0.7510** (0.2784) 0.009	1.8242* (0.7640) 0.020	9.6047*** (1.5641) 0.000	5.2452*** (1.2640) 0.000
School is in Block 2b	-0.3656*** (0.0517) 0.000	-0.0122*** (0.0035) 0.001	-0.0054 (0.0084) 0.519	-0.0193† (0.0111) 0.088
Charter school	0.0033 (0.0607) 0.957	0.0072 (0.0050) 0.154	-0.0014 (0.0124) 0.912	0.0133 (0.0166) 0.428
Total number of students enrolled	0.0001*** (0.0000) 0.000	-0.0000 (0.0000) 0.649	-0.0000* (0.0000) 0.013	-0.0000* (0.0000) 0.040
Locale type suburb	0.0055 (0.0547) 0.920	-0.0034 (0.0039) 0.383	0.0112 (0.0094) 0.238	0.0073 (0.0123) 0.555
Locale type rural or town	0.0845 (0.0629) 0.184	0.0019 (0.0047) 0.695	0.0133 (0.0111) 0.234	0.0140 (0.0148) 0.350
Proportion of students who are female	-0.1978 (0.3222) 0.542	-0.0403† (0.0208) 0.058	-0.3292*** (0.0537) 0.000	-0.4483*** (0.0653) 0.000
Proportion of students who are Black	-0.0573 (0.1618) 0.724	0.0056 (0.0110) 0.611	0.0434 (0.0264) 0.106	0.0568 (0.0347) 0.107
Proportion of students who are Hispanic	-0.1366 (0.1328) 0.308	-0.0087 (0.0081) 0.287	0.0214 (0.0189) 0.262	0.0318 (0.0244) 0.198
Within-state standardized proportion of students who are economically disadvantaged	-0.0305 (0.0317) 0.340	0.0021 (0.0022) 0.343	0.0042 (0.0052) 0.420	0.0071 (0.0069) 0.308
Constant	0.1062 (0.1993) 0.596	0.0281* (0.0124) 0.027	0.1269*** (0.0340) 0.000	0.1916*** (0.0426) 0.000
Observations	72	72	72	72
R-squared	0.714	0.472	0.778	0.715

***$p < .001$, **$p < .01$, *$p < .05$, †$p < .10$

Note. Each cell contains the coefficient, the standard error (in parentheses), and the *p* value.

Exhibit D-9. Teacher Report on the Percentage of Their AP CSP Students Who Took the AP CSP Exam, 2019–2022



Note. The two bar segments without percentages are both 3%. Percentages might not sum to 100% due to rounding. Data are from surveys administered by the evaluation to selected AP CSP teachers in 2020, 2021, 2022, and 2023 (see Chapters 3 and 4). The 2019 AP CSP exam percentages were provided by 2020 survey respondents who reported having taught AP CSP during the 2018–19 school year. Although there was some overlap from year to year in who responded to the survey, the samples were far from the same across the years, so this chart is not based on true longitudinal data. Moreover, the teachers who responded to the survey cannot be assumed to be representative of those included in the AP CSP exam impact analyses presented in Chapters 5, 6, and 7. This exhibit is provided as "a rough look" only.

Appendix E. Supplementary Exhibits for Chapter 6

Exhibit E-1. Within-Block Statistics on 2019–20 Enrollment in Grades 9–12, All Blocks in the Year 2 Analysis Sample

Subsample/block	Lone Star schools	Comparison schools	*p* value	Hedges' *g*
Subsample 1	1,526 (1,207) 13	1,574 (1,125) 91	0.886	-0.042
Subsample 2, Block 2a	1,273 (1,071) 16	1,254 (859) 64	0.941	0.021
Subsample 2, Block 2b	1,324 (1,385) 5	1,234 (1,164) 30	0.878	0.073

Note. In the first two columns, each cell contains the mean, the standard deviation (in parentheses), and the number of schools. The *p* values in the third column are from *t* tests.

Exhibit E-2. Within-Block Baseline Statistics for the Year 2, Subsample 2 Analysis Sample

Measure	Lone Star schools	Comparison schools	*p* value	Hedges' *g*
Block 2a (16 Lone Star schools and 64 comparison schools)				
Average percentage (across all 38 AP subject area exams) of 2020 exam takers earning a 3, 4, or 5	32.61% (16.18)	34.59% (16.63)	0.669	-0.119
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 3, 4, or 5 in 2020	0.28% (0.26)	0.30% (0.30)	0.780	-0.078
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 1 or 2 in 2020	0.51% (0.39)	0.48% (0.36)	0.788	0.075
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 taking a 2020 AP exam	0.78% (0.55)	0.78% (0.58)	0.976	0.008
Block 2b (5 Lone Star schools and 30 comparison schools)				
Average percentage (across all 38 AP subject area exams) of 2020 exam takers earning a 3, 4, or 5	35.87% (17.47)	38.98% (12.68)	0.633	-0.228
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 3, 4, or 5 in 2020	0.28% (0.31)	0.28% (0.27)	0.992	-0.005
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 1 or 2 in 2020	0.33% (0.23)	0.34% (0.19)	0.922	-0.046
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 taking a 2020 AP exam	0.61% (0.53)	0.62% (0.45)	0.962	-0.023

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from *t* tests.

Exhibit E-3. 2021–22 School Characteristics for the Year 2, Subsample 1 Analysis Sample, Continuous Variables

Variable	Lone Star schools (*n* = 13)	Comparison schools (*n* = 91)	*p* value	Hedges' *g*
Total number of students enrolled	1,628 (1,098)	1,669 (1,098)	0.901	-0.037
Percentage of students who are Black	12.3% (10.5)	11.2% (14.6)	0.797	0.076
Percentage of students who are Hispanic	65.5% (24.9)	32.6% (26.5)	0.000	1.241
Percentage of students who are female	47.1% (3.8)	48.9% (3.5)	0.088	-0.507
Within-state standardized percentage of students who are economically disadvantaged (missing values = 0)	-0.05 (0.67)	-0.30 (1.07)	0.406	0.246

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from *t* tests.

Exhibit E-4. 2021–22 School Characteristics for the Year 2, Subsample 1 Analysis Sample, Dichotomous Variables

Variable	Lone Star schools (*n* = 13)	Comparison schools (*n* = 91)	*p* value	Cox effect size
Charter school	38.5%	11.0%	0.014	0.976
Locale type city	53.8%	24.2%	0.033	0.780
Locale type suburb	30.8%	53.8%	0.130	-0.581
Locale type rural or town	15.4%	22.0%	0.589	-0.263
Missing percentage of students who are economically disadvantaged	0.0%	2.2%	n/a	n/a

Note. The *p* values in the third column are from logistic regression. See What Works Clearinghouse (n.d.) for the formula for the Cox effect size.

Exhibit E-5. 2021–22 School Characteristics for the Year 2, Subsample 2 Analysis Sample, Continuous Variables

Variable	Lone Star schools (*n* = 21)	Comparison schools (*n* = 94)	*p* value	Hedges' *g*
Total number of students enrolled	1,446 (964)	1,356 (895)	0.681	0.099
Percentage of students who are Black	19.6% (15.5)	15.0% (20.9)	0.341	0.229
Percentage of students who are Hispanic	62.1% (25.5)	51.9% (34.4)	0.203	0.307
Percentage of students who are female	45.1% (10.6)	48.9% (4.9)	0.012	-0.610
Within-state standardized percentage of students who are economically disadvantaged	0.30 (0.84)	0.23 (0.88)	0.742	0.079

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from *t* tests.

Exhibit E-6. 2021–22 School Characteristics for the Year 2, Subsample 2 Analysis Sample, Dichotomous Variables

Variable	Lone Star schools (*n* = 21)	Comparison schools (*n* = 94)	*p* value	Cox effect size
Charter school	52.4%	17.0%	0.001	1.011
Locale type city	71.4%	44.7%	0.032	0.680
Locale type suburb	28.6%	26.6%	0.854	0.060
Locale type rural or town	0.0%	28.7%	n/a	n/a
Missing percentage of students who are economically disadvantaged	0.0%	1.1%	n/a	n/a

Note. The *p* values in the third column are from logistic regression. See What Works Clearinghouse (n.d.) for the formula for the Cox effect size.

Exhibit E-7. Full Model Regression Results for the Year 2, Subsample 1 Analyses

Variable	Proportion of 2022 AP CSP exam takers earning a 3, 4, or 5 on the exam	Proportion of students in grades 9–12 earning a 3, 4, or 5 on the 2022 AP CSP exam	Proportion of students in grades 9–12 earning a 1 or 2 on the 2022 AP CSP exam	Proportion of students in grades 9–12 taking the 2022 AP CSP exam
Lone Star school	-0.0911 (0.0670) 0.177	-0.0060* (0.0025) 0.018	-0.0054 (0.0039) 0.165	-0.0124** (0.0043) 0.005
Baseline measure of the outcome	0.3954*** (0.1134) 0.001	0.3858*** (0.1059) 0.000	1.0269*** (0.1156) 0.000	0.8981*** (0.0940) 0.000
Charter school	-0.0342 (0.0722) 0.637	0.0081** (0.0028) 0.005	0.0138** (0.0046) 0.004	0.0208*** (0.0052) 0.000
Total number of students enrolled	0.0000 (0.0000) 0.307	-0.0000** (0.0000) 0.004	-0.0000 (0.0000) 0.768	-0.0000 (0.0000) 0.359
Locale type suburb	-0.0610 (0.0523) 0.246	-0.0020 (0.0019) 0.301	0.0005 (0.0031) 0.884	-0.0004 (0.0034) 0.914
Locale type rural or town	-0.0292 (0.0660) 0.659	-0.0005 (0.0024) 0.841	0.0003 (0.0039) 0.936	0.0004 (0.0043) 0.922
Proportion of students who are female	-0.5452 (0.5619) 0.334	-0.1173*** (0.0212) 0.000	-0.0162 (0.0327) 0.622	-0.1351*** (0.0364) 0.000
Proportion of students who are Black	-0.3423† (0.1860) 0.069	-0.0062 (0.0069) 0.373	-0.0132 (0.0105) 0.212	-0.0150 (0.0118) 0.205
Proportion of students who are Hispanic	-0.3197** (0.1004) 0.002	0.0009 (0.0035) 0.790	0.0139* (0.0055) 0.014	0.0200** (0.0060) 0.001
Within-state standardized proportion of students who are economically disadvantaged	-0.0462* (0.0226) 0.044	-0.0024** (0.0009) 0.007	0.0010 (0.0013) 0.437	-0.0008 (0.0015) 0.611
Missing proportion of students who are economically disadvantaged	-0.0432 (0.1680) 0.798	0.0064 (0.0063) 0.317	0.0148 (0.0098) 0.134	0.0218* (0.0109) 0.048
Constant	0.7628* (0.2927) 0.011	0.0701*** (0.0107) 0.000	0.0068 (0.0160) 0.671	0.0668*** (0.0181) 0.000
Observations	104	104	104	104
R-squared	0.520	0.627	0.719	0.786

***p < .001, **p < .01, *p < .05, †p < .10

Note. Each cell contains the coefficient, the standard error (in parentheses), and the p value.

Exhibit E-8. Full Model Regression Results for the Year 2, Subsample 2 Analyses

Variable	Proportion of 2022 AP CSP exam takers earning a 3, 4, or 5 on the exam	Proportion of students in grades 9–12 earning a 3, 4, or 5 on the 2022 AP CSP exam	Proportion of students in grades 9–12 earning a 1 or 2 on the 2022 AP CSP exam	Proportion of students in grades 9–12 taking the 2022 AP CSP exam
Lone Star school	-0.0640 (0.0479) 0.184	-0.0067† (0.0036) 0.066	-0.0064 (0.0062) 0.302	-0.0123 (0.0085) 0.154
Baseline measure of the outcome	0.5486** (0.1653) 0.001	0.4383 (0.6800) 0.521	4.2631*** (0.8856) 0.000	3.0888*** (0.7884) 0.000
School is in Block 2b	-0.4334*** (0.0396) 0.000	-0.0100*** (0.0029) 0.001	-0.0060 (0.0049) 0.223	-0.0152* (0.0067) 0.026
Charter school	-0.0119 (0.0517) 0.818	0.0106* (0.0042) 0.014	0.0128† (0.0066) 0.053	0.0181† (0.0097) 0.065
Total number of students enrolled	0.0001* (0.0000) 0.011	-0.0000 (0.0000) 0.604	-0.0000 (0.0000) 0.222	-0.0000 (0.0000) 0.310
Locale type suburb	0.0492 (0.0427) 0.252	0.0042 (0.0031) 0.179	-0.0018 (0.0052) 0.730	0.0021 (0.0072) 0.775
Locale type rural or town	-0.0046 (0.0544) 0.932	0.0010 (0.0042) 0.818	0.0058 (0.0069) 0.400	0.0092 (0.0096) 0.338
Proportion of students who are female	0.0323 (0.2940) 0.913	-0.0061 (0.0206) 0.767	-0.2097*** (0.0355) 0.000	-0.2526*** (0.0461) 0.000
Proportion of students who are Black	-0.1507 (0.1278) 0.241	0.0001 (0.0092) 0.994	0.0282† (0.0153) 0.069	0.0324 (0.0212) 0.130
Proportion of students who are Hispanic	-0.2029* (0.0935) 0.032	-0.0023 (0.0064) 0.717	0.0146 (0.0109) 0.184	0.0190 (0.0147) 0.200
Within-state standardized proportion of students who are economically disadvantaged	-0.0482* (0.0237) 0.044	0.0009 (0.0018) 0.598	0.0053† (0.0029) 0.067	0.0076† (0.0040) 0.059
Missing proportion of students who are economically disadvantaged	0.2515 (0.1922) 0.194	0.0249† (0.0142) 0.082	0.0027 (0.0231) 0.908	0.0134 (0.0318) 0.674
Constant	0.2265 (0.1496) 0.133	0.0110 (0.0108) 0.313	0.0867*** (0.0206) 0.000	0.1082*** (0.0266) 0.000
Observations	115	115	115	115
R-squared	0.635	0.293	0.618	0.548

***$p < .001$, **$p < .01$, *$p < .05$, †$p < .10$

Note. Each cell contains the coefficient, the standard error (in parentheses), and the p value.

Exhibit E-9. Regression Results for Three Models Predicting the Participation Outcome (in 0-to-1 Proportion Metric) for Year 2, Subsample 1

Variable	Model 1: Lone Star indicator only	Model 2: Lone Star indicator and baseline measure	Model 3: Full model with all covariates
Lone Star school	0.0018 (0.0076) 0.816	0.0016 (0.0045) 0.725	-0.0124** (0.0043) 0.005
Baseline measure of the outcome		1.1247*** (0.0820) 0.000	0.8981*** (0.0940) 0.000
Charter school			0.0208*** (0.0052) 0.000
Total number of students enrolled			-0.0000 (0.0000) 0.359
Locale type suburb			-0.0004 (0.0034) 0.914
Locale type rural or town			0.0004 (0.0043) 0.922
Proportion of students who are female			-0.1351*** (0.0364) 0.000
Proportion of students who are Black			-0.0150 (0.0118) 0.205
Proportion of students who are Hispanic			0.0200** (0.0060) 0.001
Within-state standardized proportion of students who are economically disadvantaged			-0.0008 (0.0015) 0.611
Missing proportion of students who are economically disadvantaged			0.0218* (0.0109) 0.048
Constant	0.0236*** (0.0027) 0.000	0.0015 (0.0023) 0.511	0.0668*** (0.0181) 0.000
Observations	104	104	104
R-squared	0.001	0.651	0.786

***p < .001, **p < .01, *p < .05, †p < .10

Note. Each cell contains the coefficient, the standard error (in parentheses), and the p value.

Exhibit E-10. Regression Results for Four Models Predicting the Participation Outcome (in 0-to-1 Proportion Metric) for Year 2, Subsample 2

Variable	Model 1: Lone Star indicator only	Model 2: Lone Star indicator and baseline measure	Model 3: Lone Star indicator, baseline measure, block	Model 4: Full model with all covariates
Lone Star school	0.0064 (0.0103) 0.536	0.0058 (0.0087) 0.507	0.0040 (0.0084) 0.638	-0.0123 (0.0085) 0.154
Baseline measure of the outcome		4.1958*** (0.6250) 0.000	3.9292*** (0.6052) 0.000	3.0888*** (0.7884) 0.000
School is in Block 2b			-0.0232** (0.0071) 0.001	-0.0152* (0.0067) 0.026
Charter school				0.0181† (0.0097) 0.065
Total number of students enrolled				-0.0000 (0.0000) 0.310
Locale type suburb				0.0021 (0.0072) 0.775
Locale type rural or town				0.0092 (0.0096) 0.338
Proportion of students who are female				-0.2526*** (0.0461) 0.000
Proportion of students who are Black				0.0324 (0.0212) 0.130
Proportion of students who are Hispanic				0.0190 (0.0147) 0.200
Within-state standardized proportion of students who are economically disadvantaged				0.0076† (0.0040) 0.059
Missing proportion of students who are economically disadvantaged				0.0134 (0.0318) 0.674
Constant	0.0194*** (0.0044) 0.000	-0.0112† (0.0059) 0.060	-0.0019 (0.0063) 0.768	0.1082*** (0.0266) 0.000
Observations	115	115	115	115
R-squared	0.003	0.289	0.352	0.548

***p < .001, **p < .01, *p < .05, †p < .10

Note. Each cell contains the coefficient, the standard error (in parentheses), and the p value.

Appendix F. Supplementary Exhibits for Chapter 7

Exhibit F-1. Within-Block Statistics on 2019–20 Enrollment in Grades 9–12, All Blocks in the Year 3 Analysis Sample

Subsample/block	Lone Star schools	Comparison schools	*p* value	Hedges' *g*
Subsample 1	1,644 (1,180) 12	1,698 (1,103) 96	0.874	-0.048
Subsample 2, Block 2a	1,173 (958) 14	1,282 (861) 84	0.668	-0.123
Subsample 2, Block 2b	1,616 (1,548) 6	1,529 (738) 36	0.824	0.097

Note. In the first two columns, each cell contains the mean, the standard deviation (in parentheses), and the number of schools. The *p* values in the third column are from *t* tests.

Exhibit F-2. Within-Block Baseline Statistics for the Year 3, Subsample 2 Analysis Sample

Measure	Lone Star schools	Comparison schools	*p* value	Hedges' *g*
Block 2a (14 Lone Star schools and 84 comparison schools)				
Average percentage (across all 38 AP subject area exams) of 2020 exam takers earning a 3, 4, or 5	37.78% (16.11)	39.62% (17.34)	0.712	-0.106
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 3, 4, or 5 in 2020	0.31% (0.27)	0.31% (0.27)	0.954	-0.017
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 1 or 2 in 2020	0.40% (0.17)	0.40% (0.24)	0.996	-0.002
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 taking a 2020 AP exam	0.71% (0.41)	0.72% (0.43)	0.969	-0.011
Block 2b (6 Lone Star schools and 36 comparison schools)				
Average percentage (across all 38 AP subject area exams) of 2020 exam takers earning a 3, 4, or 5	27.14% (12.08)	27.42% (11.41)	0.956	-0.024
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 3, 4, or 5 in 2020	0.25% (0.27)	0.26% (0.28)	0.969	-0.017
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 earning a 1 or 2 in 2020	0.61% (0.64)	0.74% (0.94)	0.745	-0.142
Average percentage (across all 38 AP subject area exams) of students in grades 9–12 taking a 2020 AP exam	0.86% (0.84)	1.00% (1.18)	0.789	-0.116

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from *t* tests.

Exhibit F-3. 2022–23 School Characteristics for the Year 3, Subsample 1 Analysis Sample, Continuous Variables

Variable	Lone Star schools (*n* = 12)	Comparison schools (*n* = 96)	*p* value	Hedges' *g*
Total number of students enrolled	1,741 (1,104)	1,748 (1,085)	0.984	-0.006
Percentage of students who are Black	11.5% (10.4)	11.8% (14.0)	0.947	-0.020
Percentage of students who are Hispanic	66.7% (25.5)	33.4% (24.8)	0.000	1.328
Percentage of students who are female	47.6% (1.5)	49.0% (4.1)	0.226	-0.371
Within-state standardized percentage of students who are economically disadvantaged (missing values = 0)	0.07 (0.85)	-0.39 (0.96)	0.115	0.483

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from *t* tests.

Exhibit F-4. 2022–23 School Characteristics for the Year 3, Subsample 1 Analysis Sample, Dichotomous Variables

Variable	Lone Star schools (*n* = 12)	Comparison schools (*n* = 96)	*p* value	Cox effect size
Charter school	33.3%	5.2%	0.004	1.329
Locale type city	58.3%	21.9%	0.011	0.968
Locale type suburb	25.0%	53.1%	0.079	-0.736
Locale type rural or town	16.7%	25.0%	0.528	-0.307
Missing percentage of students who are economically disadvantaged	0.0%	2.1%	n/a	n/a

Note. The *p* values in the third column are from logistic regression. See What Works Clearinghouse (n.d.) for the formula for the Cox effect size.

Exhibit F-5. 2022–23 School Characteristics for the Year 3, Subsample 2 Analysis Sample, Continuous Variables

Variable	Lone Star schools (*n* = 20)	Comparison schools (*n* = 120)	*p* value	Hedges' *g*
Total number of students enrolled	1,497 (963)	1,427 (774)	0.717	0.087
Percentage of students who are Black	18.4% (15.1)	13.7% (19.5)	0.302	0.249
Percentage of students who are Hispanic	61.9% (26.8)	56.5% (34.2)	0.504	0.161
Percentage of students who are female	45.2% (10.8)	48.9% (4.4)	0.007	-0.655
Within-state standardized percentage of students who are economically disadvantaged	0.11 (0.97)	0.29 (0.92)	0.437	-0.187

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from *t* tests.

Exhibit F-6. 2022–23 School Characteristics for the Year 3, Subsample 2 Analysis Sample, Dichotomous Variables

Variable	Lone Star schools (*n* = 20)	Comparison schools (*n* = 120)	*p* value	Cox effect size
Charter school	55.0%	8.3%	0.000	1.566
Locale type city	70.0%	40.8%	0.020	0.734
Locale type suburb	30.0%	29.2%	0.940	0.024
Locale type rural or town	0.0%	30.0%	n/a	n/a
Missing percentage of students who are economically disadvantaged	0.0%	1.7%	n/a	n/a

Note. The *p* values in the third column are from logistic regression. See What Works Clearinghouse (n.d.) for the formula for the Cox effect size.

Exhibit F-7. Full Model Regression Results for the Year 3, Subsample 1 Analyses

Variable	Proportion of 2023 AP CSP exam takers earning a 3, 4, or 5 on the exam	Proportion of students in grades 9–12 earning a 3, 4, or 5 on the 2023 AP CSP exam	Proportion of students in grades 9–12 earning a 1 or 2 on the 2023 AP CSP exam	Proportion of students in grades 9–12 taking the 2023 AP CSP exam
Lone Star school	-0.0954 (0.0663) 0.154	-0.0075* (0.0037) 0.046	-0.0051 (0.0031) 0.108	-0.0129* (0.0054) 0.020
Baseline measure of the outcome	0.4181*** (0.1193) 0.001	0.8164*** (0.1618) 0.000	0.9941** (0.3304) 0.003	1.0391*** (0.2017) 0.000
Charter school	0.0427 (0.0778) 0.585	0.0090* (0.0044) 0.045	0.0100** (0.0037) 0.008	0.0177** (0.0065) 0.008
Total number of students enrolled	-0.0000 (0.0000) 0.883	-0.0000† (0.0000) 0.095	-0.0000 (0.0000) 0.154	-0.0000 (0.0000) 0.153
Locale type suburb	0.0016 (0.0475) 0.973	-0.0044 (0.0027) 0.105	-0.0027 (0.0023) 0.233	-0.0067† (0.0040) 0.094
Locale type rural or town	0.0430 (0.0568) 0.451	-0.0014 (0.0032) 0.657	-0.0009 (0.0027) 0.752	-0.0023 (0.0047) 0.627
Proportion of students who are female	0.4846 (0.4651) 0.300	-0.0780** (0.0265) 0.004	-0.0572* (0.0224) 0.012	-0.1411*** (0.0390) 0.000
Proportion of students who are Black	-0.2148 (0.1596) 0.181	0.0010 (0.0088) 0.908	0.0036 (0.0073) 0.621	0.0064 (0.0128) 0.616
Proportion of students who are Hispanic	-0.1270 (0.1000) 0.207	0.0025 (0.0056) 0.660	0.0039 (0.0045) 0.394	0.0084 (0.0081) 0.301
Within-state standardized proportion of students who are economically disadvantaged	-0.0819*** (0.0229) 0.001	-0.0018 (0.0013) 0.163	0.0029** (0.0011) 0.008	0.0012 (0.0019) 0.528
Missing proportion of students who are economically disadvantaged	0.0576 (0.1367) 0.675	0.0035 (0.0077) 0.653	0.0000 (0.0065) 0.999	0.0033 (0.0112) 0.773
Constant	0.1694 (0.2608) 0.517	0.0485*** (0.0136) 0.001	0.0354** (0.0113) 0.002	0.0814*** (0.0198) 0.000
Observations	108	108	108	108
R-squared	0.484	0.531	0.418	0.539

***p < .001, **p < .01, *p < .05, †p < .10

Note. Each cell contains the coefficient, the standard error (in parentheses), and the p value.

Exhibit F-8. Full Model Regression Results for the Year 3, Subsample 2 Analyses

Variable	Proportion of 2023 AP CSP exam takers earning a 3, 4, or 5 on the exam	Proportion of students in grades 9–12 earning a 3, 4, or 5 on the 2023 AP CSP exam	Proportion of students in grades 9–12 earning a 1 or 2 on the 2023 AP CSP exam	Proportion of students in grades 9–12 taking the 2023 AP CSP exam
Lone Star school	0.0545 (0.0532) 0.308	0.0033 (0.0031) 0.294	-0.0102* (0.0048) 0.034	-0.0070 (0.0064) 0.274
Baseline measure of the outcome	0.5756*** (0.1395) 0.000	0.9557* (0.4412) 0.032	0.3594 (0.3033) 0.238	0.5295† (0.3060) 0.086
School is in Block 2b	-0.2824*** (0.0367) 0.000	-0.0070** (0.0021) 0.001	-0.0138*** (0.0033) 0.000	-0.0213*** (0.0044) 0.000
Charter school	-0.1067† (0.0597) 0.076	0.0052 (0.0035) 0.139	0.0216*** (0.0052) 0.000	0.0274*** (0.0069) 0.000
Total number of students enrolled	-0.0000 (0.0000) 0.422	-0.0000 (0.0000) 0.131	-0.0000 (0.0000) 0.611	-0.0000 (0.0000) 0.275
Locale type suburb	0.0192 (0.0408) 0.638	-0.0004 (0.0024) 0.879	0.0002 (0.0037) 0.955	-0.0003 (0.0049) 0.944
Locale type rural or town	0.0160 (0.0470) 0.734	-0.0031 (0.0028) 0.267	-0.0009 (0.0042) 0.828	-0.0043 (0.0056) 0.451
Proportion of students who are female	0.2235 (0.2850) 0.435	0.0119 (0.0176) 0.500	0.0122 (0.0259) 0.639	0.0265 (0.0351) 0.451
Proportion of students who are Black	-0.1931 (0.1312) 0.144	-0.0085 (0.0074) 0.252	0.0195† (0.0111) 0.082	0.0096 (0.0148) 0.517
Proportion of students who are Hispanic	-0.1933* (0.0904) 0.034	-0.0123* (0.0047) 0.011	0.0037 (0.0073) 0.613	-0.0105 (0.0094) 0.267
Within-state standardized proportion of students who are economically disadvantaged	-0.0301 (0.0224) 0.182	-0.0012 (0.0014) 0.392	0.0025 (0.0020) 0.221	0.0012 (0.0027) 0.654
Missing proportion of students who are economically disadvantaged	0.1534 (0.1372) 0.265	-0.0079 (0.0081) 0.335	-0.0156 (0.0123) 0.206	-0.0226 (0.0164) 0.170
Constant	0.2257 (0.1565) 0.152	0.0118 (0.0087) 0.179	0.0008 (0.0135) 0.950	0.0133 (0.0179) 0.459
Observations	140	140	140	140
R-squared	0.649	0.397	0.345	0.390

***p < .001, **p < .01, *p < .05, †p < .10

Note. Each cell contains the coefficient, the standard error (in parentheses), and the p value.

Appendix G. Supplementary Exhibits for Addendum to Chapters 5–7

Exhibit G-1. School Characteristics for the Robustness Check Analysis Sample, Continuous Variables

Variable	Lone Star schools (*n* = 12)	Comparison schools (*n* = 84)	*p* value	Hedges' *g*
2020–21				
Total number of students enrolled	1,718 (1,075)	1,977 (998)	0.406	-0.255
Percentage of students who are Black	11.3% (10.2)	12.4% (10.0)	0.742	-0.101
Percentage of students who are Hispanic	65.6% (25.7)	50.9% (24.4)	0.056	0.592
Percentage of students who are female	48.2% (1.8)	50.2% (8.4)	0.419	-0.249
Percentage of students who are economically disadvantaged	66.5% (18.1)	48.5% (26.2)	0.023	0.706
2021–22				
Total number of students enrolled	1,734 (1,076)	1,994 (1,024)	0.416	-0.250
Percentage of students who are Black	11.5% (10.5)	12.5% (10.1)	0.759	-0.094
Percentage of students who are Hispanic	66.0% (26.0)	51.5% (24.3)	0.057	0.589
Percentage of students who are female	48.0% (2.4)	50.2% (8.5)	0.366	-0.278
Percentage of students who are economically disadvantaged	64.4% (16.7)	49.9% (25.7)	0.062	0.579
2022–23				
Total number of students enrolled	1,741 (1,104)	1,985 (1,027)	0.449	-0.233
Percentage of students who are Black	11.5% (10.4)	12.6% (10.2)	0.730	-0.106
Percentage of students who are Hispanic	66.7% (25.5)	51.8% (24.2)	0.051	0.606
Percentage of students who are female	47.6% (1.5)	50.1% (8.5)	0.302	-0.318
Percentage of students who are economically disadvantaged (missing values imputed at the mean)	70.2% (18.0)	53.7% (23.6)	0.022	0.712

Note. In the first two columns, each cell contains the mean and (in parentheses) the standard deviation. The *p* values in the third column are from *t* tests.

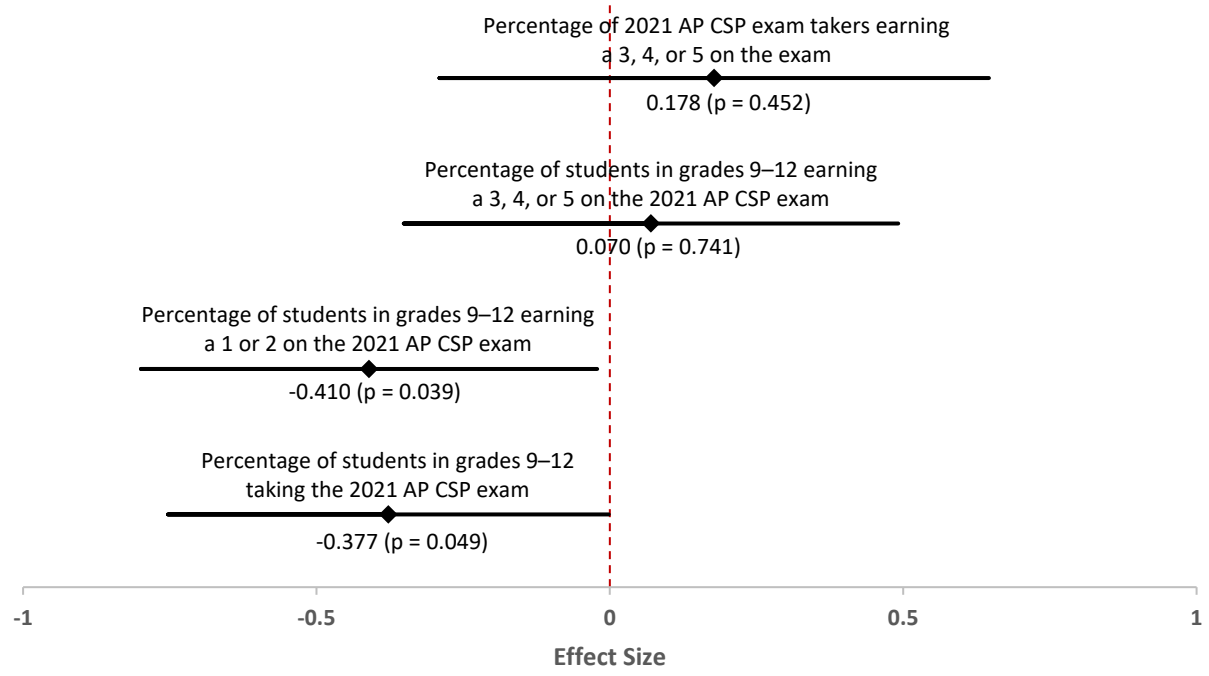
Exhibit G-2. School Characteristics for the Robustness Check Analysis Sample, Dichotomous Variables

Variable	Lone Star schools (*n* = 12)	Comparison schools (*n* = 84)	*p* value	Cox effect size
2020–21 and 2021–22				
Charter school	33.3%	15.5%	0.141	0.604
Locale type city	58.3%	42.9%	0.318	0.375
Locale type suburb	25.0%	36.9%	0.424	-0.338
Locale type rural or town	16.7%	20.2%	0.772	-0.143
Missing percentage of students who are economically disadvantaged	0.0%	0.0%	n/a	n/a
2022–23				
Charter school	33.3%	15.5%	0.141	0.604
Locale type city	58.3%	41.7%	0.282	0.405
Locale type suburb	25.0%	38.1%	0.383	-0.369
Locale type rural or town	16.7%	20.2%	0.772	-0.143
Missing percentage of students who are economically disadvantaged	0.0%	7.1%	n/a	n/a

Note. The *p* values in the third column are from logistic regression. See What Works Clearinghouse (n.d.) for the formula for the Cox effect size.

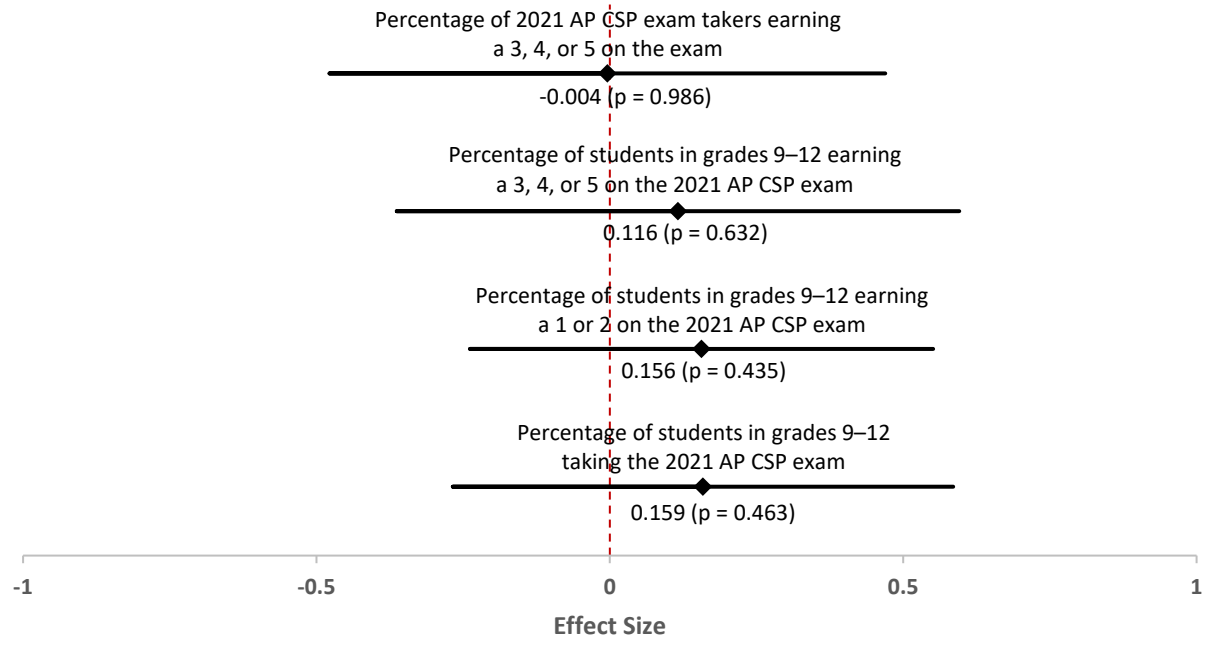
Exhibit G-3. Effect Size Charts for the Year 1 Analyses

a. Robustness check



Note. *N* = 96 (12 Lone Star schools and 84 comparison schools).

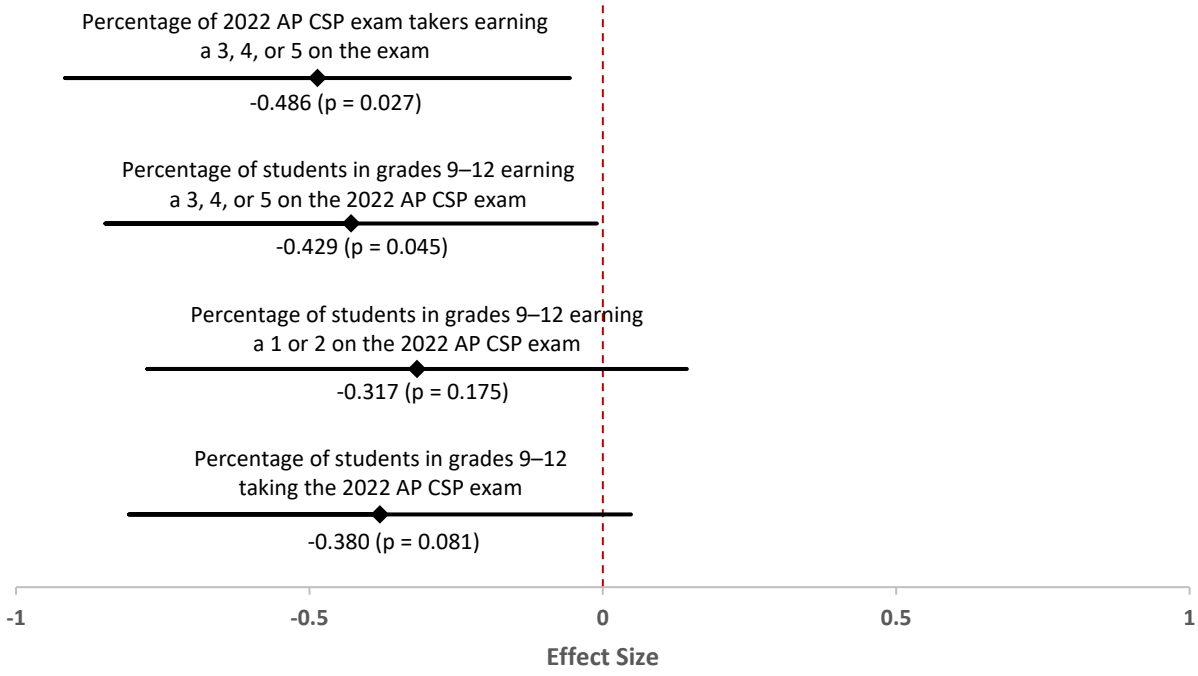
b. Main analysis (Subsample 1) from Chapter 5



Note. *N* = 143 (13 Lone Star schools and 130 comparison schools).

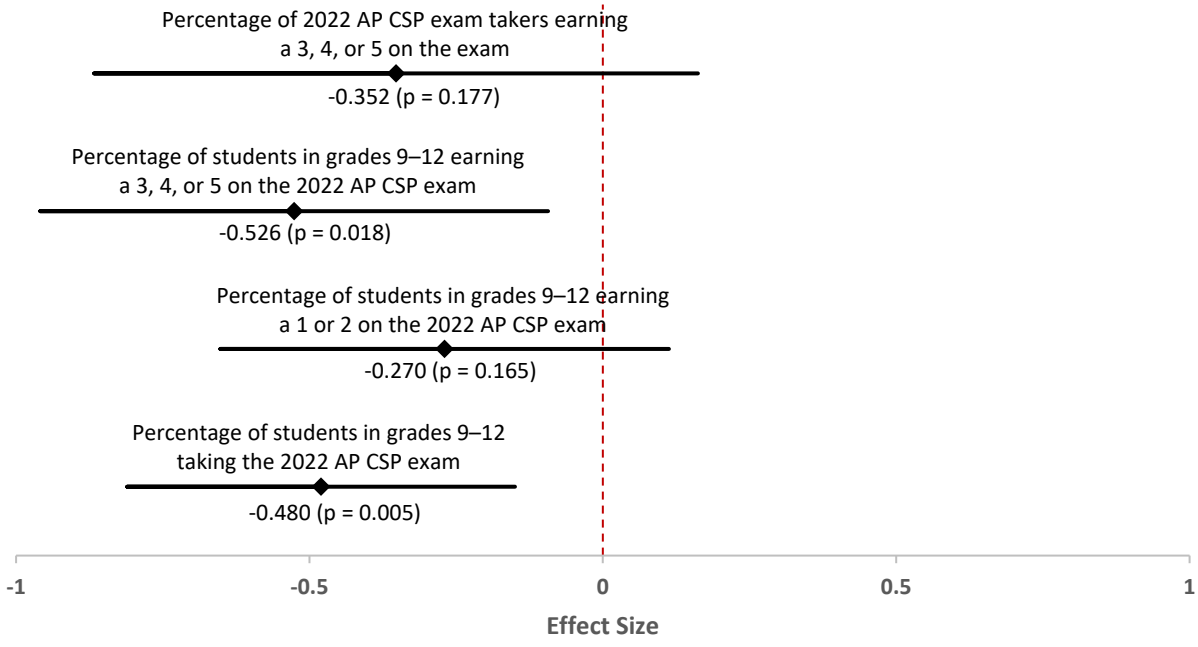
Exhibit G-4. Effect Size Charts for the Year 2 Analyses

a. Robustness check



Note. *N* = 96 (12 Lone Star schools and 84 comparison schools).

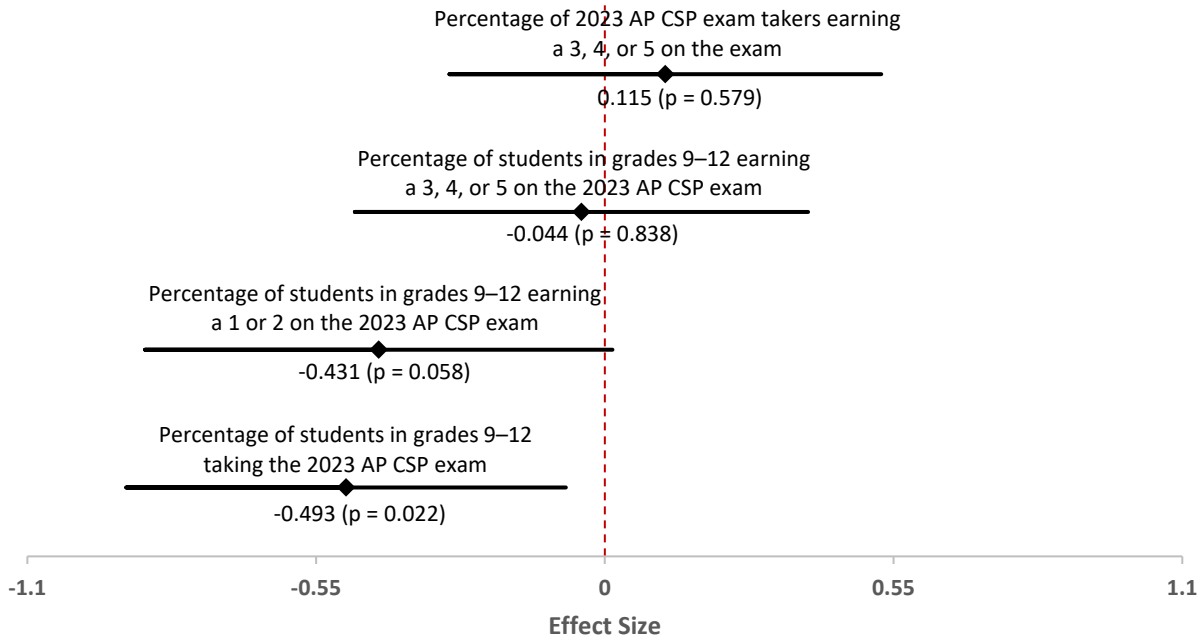
b. Main analysis (Subsample 1) from Chapter 6



Note. *N* = 104 (13 Lone Star schools and 91 comparison schools).

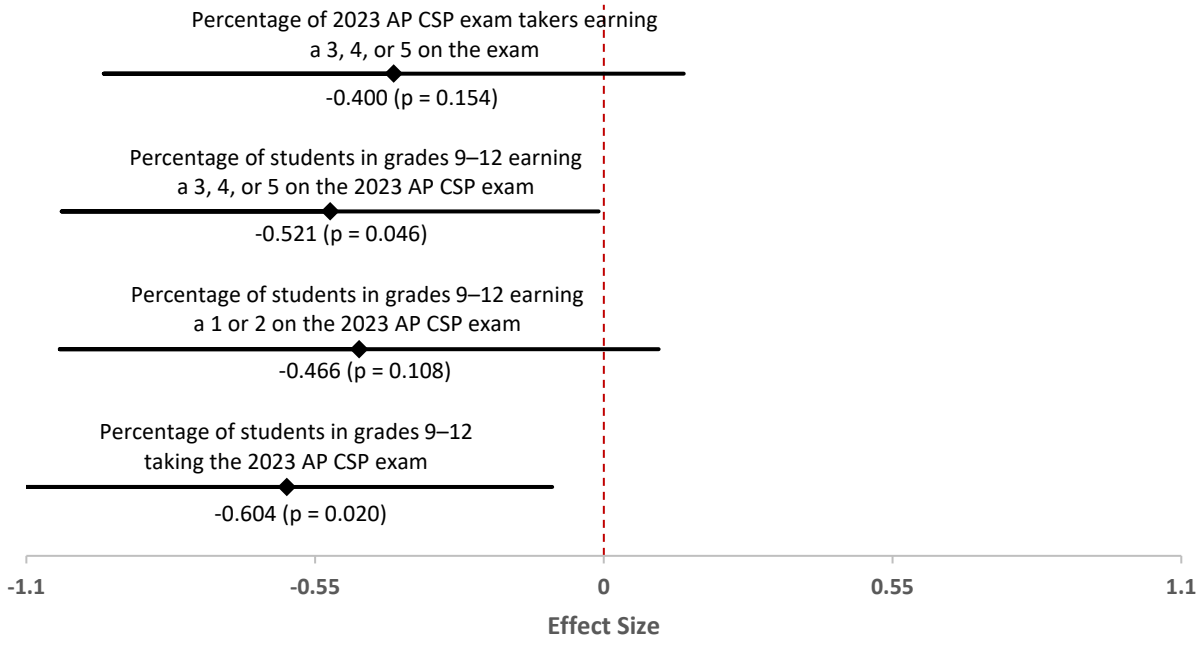
Exhibit G-5. Effect Size Charts for the Year 3 Analyses

a. Robustness check



Note. *N* = 96 (12 Lone Star schools and 84 comparison schools).

b. Main analysis (Subsample 1) from Chapter 7



Note. *N* = 108 (12 Lone Star schools and 96 comparison schools).

Exhibit G-6. Outcome Statistics for Robustness Check Sample

Measure	Lone Star schools (unadj.) (n = 12)	Comparison schools (unadj.) (n = 84)	Adj. mean difference (θ_{Trt})	p value	Hedges' g
Year 1					
Percentage of 2021 AP CSP exam takers earning a 3, 4, or 5 on the exam	54.23% (21.32)	60.34% (26.58)	4.66% (6.18)	0.452	0.178
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2021 AP CSP exam	0.96% (0.74)	1.17% (1.38)	0.09% (0.28)	0.741	0.070
Percentage of students in grades 9–12 earning a 1 or 2 on the 2021 AP CSP exam	0.78% (0.69)	1.35% (2.57)	-1.01% (0.48)	0.039	-0.410
Percentage of students in grades 9–12 taking the 2021 AP CSP exam	1.75% (1.16)	2.52% (3.21)	-1.16% (0.58)	0.049	-0.377
Year 2					
Percentage of 2022 AP CSP exam takers earning a 3, 4, or 5 on the exam	40.04% (16.24)	61.24% (26.72%	-12.60% (5.61)	0.027	-0.486
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2022 AP CSP exam	0.76% (0.56)	1.31% (1.23)	-0.51% (0.25)	0.045	-0.429
Percentage of students in grades 9–12 earning a 1 or 2 on the 2022 AP CSP exam	1.36% (1.18)	1.50% (3.17)	-0.96% (0.70)	0.175	-0.317
Percentage of students in grades 9–12 taking the 2022 AP CSP exam	2.12% (1.53)	2.82% (3.79)	-1.38% (0.78)	0.081	-0.380
Year 3					
Percentage of 2023 AP CSP exam takers earning a 3, 4, or 5 on the exam	49.28% 24.97%	55.81% 26.94%	3.10% (5.57)	0.579	0.115
Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2023 AP CSP exam	1.25% 1.55%	1.48% 1.47%	-0.07% (0.32)	0.838	-0.044
Percentage of students in grades 9–12 earning a 1 or 2 on the 2023 AP CSP exam	1.02% 0.73%	2.00% 4.17%	-1.71% (0.89)	0.058	-0.431
Percentage of students in grades 9–12 taking the 2023 AP CSP exam	2.27% 1.74%	3.48% 4.71%	-2.22% (0.95)	0.022	-0.493

Note. In the first two columns, each cell contains the unadjusted mean and (in parentheses) standard deviation of the outcome measure. The third column contains the regression coefficient on the Lone Star indicator from the impact analysis model (i.e., the adjusted mean difference) and (in parentheses) the standard error of the coefficient; the p value associated with the coefficient is in the fourth column. The Hedges' g value in the final column results from the inclusion of the regression coefficient as the numerator (i.e., the mean difference) in the Hedges' g formula.

Exhibit G-7. Full Model Regression Results for the Year 1 Robustness Check Analyses

Variable	Proportion of 2021 AP CSP exam takers earning a 3, 4, or 5 on the exam	Proportion of students in grades 9–12 earning a 3, 4, or 5 on the 2021 AP CSP exam	Proportion of students in grades 9–12 earning a 1 or 2 on the 2021 AP CSP exam	Proportion of students in grades 9–12 taking the 2021 AP CSP exam
Lone Star school	0.0466 (0.0618) 0.452	0.0009 (0.0028) 0.741	-0.0101* (0.0048) 0.039	-0.0116* (0.0058) 0.049
Baseline measure of the outcome	0.3504** (0.1157) 0.003	0.6505*** (0.0885) 0.000	2.4755*** (0.4161) 0.000	1.0607*** (0.1510) 0.000
Charter school	-0.1111† (0.0668) 0.100	-0.0054† (0.0032) 0.097	0.0131* (0.0054) 0.018	0.0096 (0.0067) 0.158
Total number of students enrolled	0.0000 (0.0000) 0.145	-0.0000 (0.0000) 0.184	-0.0000 (0.0000) 0.480	-0.0000 (0.0000) 0.130
Locale type suburb	-0.0380 (0.0488) 0.438	-0.0031 (0.0022) 0.171	0.0029 (0.0037) 0.432	0.0009 (0.0046) 0.853
Locale type rural or town	0.0233 (0.0588) 0.693	-0.0032 (0.0027) 0.240	0.0031 (0.0043) 0.479	0.0028 (0.0054) 0.601
Proportion of students who are female	0.2868 (0.2823) 0.313	-0.0121 (0.0130) 0.356	-0.0496* (0.0220) 0.027	-0.0538* (0.0268) 0.048
Proportion of students who are Black	-0.3748 (0.3259) 0.253	-0.0030 (0.0147) 0.839	-0.0184 (0.0244) 0.453	-0.0171 (0.0300) 0.570
Proportion of students who are Hispanic	-0.4315† (0.2212) 0.054	-0.0086 (0.0099) 0.388	-0.0059 (0.0170) 0.729	0.0008 (0.0203) 0.968
Proportion of students who are economically disadvantaged	-0.1139 (0.2049) 0.580	-0.0037 (0.0094) 0.691	0.0283† (0.0158) 0.076	0.0190 (0.0193) 0.325
Constant	0.4965* (0.1904) 0.011	0.0215** (0.0075) 0.005	0.0157 (0.0121) 0.199	0.0301† (0.0151) 0.050
Observations	96	96	96	96
R-squared	0.511	0.605	0.680	0.692

***$p < .001$, **$p < .01$, *$p < .05$, †$p < .10$

Note. Each cell contains the coefficient, the standard error (in parentheses), and the *p* value.

Exhibit G-8. Full Model Regression Results for the Year 2 Robustness Check Analyses

Variable	Proportion of 2022 AP CSP exam takers earning a 3, 4, or 5 on the exam	Proportion of students in grades 9–12 earning a 3, 4, or 5 on the 2022 AP CSP exam	Proportion of students in grades 9–12 earning a 1 or 2 on the 2022 AP CSP exam	Proportion of students in grades 9–12 taking the 2022 AP CSP exam
Lone Star school	-0.1260* (0.0561) 0.027	-0.0051* (0.0025) 0.045	-0.0096 (0.0070) 0.175	-0.0138† (0.0078) 0.081
Baseline measure of the outcome	0.3605** (0.1059) 0.001	0.4336*** (0.0811) 0.000	0.9294 (0.6189) 0.137	0.7636*** (0.2115) 0.001
Charter school	-0.0849 (0.0607) 0.166	0.0005 (0.0029) 0.859	0.0279*** (0.0080) 0.001	0.0247** (0.0092) 0.009
Total number of students enrolled	0.0000 (0.0000) 0.245	-0.0000** (0.0000) 0.003	-0.0000 (0.0000) 0.163	-0.0000† (0.0000) 0.082
Locale type suburb	-0.0827† (0.0443) 0.066	-0.0030 (0.0020) 0.135	0.0120* (0.0053) 0.027	0.0105† (0.0062) 0.092
Locale type rural or town	-0.0452 (0.0540) 0.404	-0.0017 (0.0024) 0.477	0.0185** (0.0063) 0.004	0.0194* (0.0074) 0.010
Proportion of students who are female	-0.2942 (0.2599) 0.261	-0.0168 (0.0116) 0.151	0.0312 (0.0322) 0.334	0.0117 (0.0360) 0.746
Proportion of students who are Black	-0.3411 (0.3096) 0.274	0.0001 (0.0137) 0.997	-0.0318 (0.0374) 0.398	-0.0359 (0.0425) 0.401
Proportion of students who are Hispanic	-0.4327* (0.1942) 0.029	-0.0042 (0.0086) 0.624	0.0073 (0.0246) 0.766	0.0001 (0.0271) 0.998
Proportion of students who are economically disadvantaged	-0.1808 (0.1878) 0.338	-0.0054 (0.0086) 0.532	0.0326 (0.0235) 0.169	0.0355 (0.0271) 0.194
Constant	0.8716*** (0.1758) 0.000	0.0287*** (0.0067) 0.000	-0.0259 (0.0177) 0.147	-0.0047 (0.0207) 0.820
Observations	96	96	96	96
R-squared	0.612	0.618	0.553	0.601

***$p < .001$, **$p < .01$, *$p < .05$, †$p < .10$

Note. Each cell contains the coefficient, the standard error (in parentheses), and the *p* value.

Exhibit G-9. Full Model Regression Results for the Year 3 Robustness Check Analyses

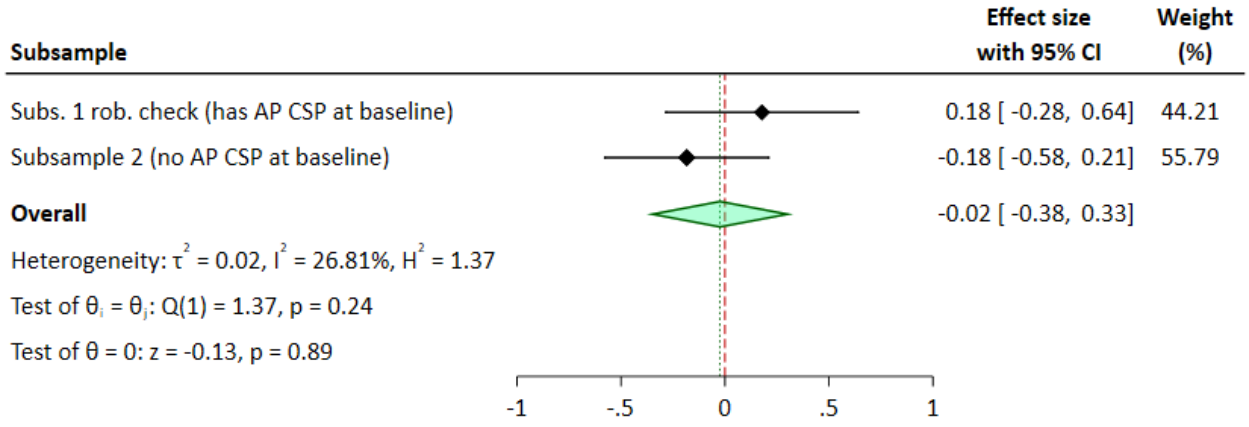
Variable	Proportion of 2023 AP CSP exam takers earning a 3, 4, or 5 on the exam	Proportion of students in grades 9–12 earning a 3, 4, or 5 on the 2023 AP CSP exam	Proportion of students in grades 9–12 earning a 1 or 2 on the 2023 AP CSP exam	Proportion of students in grades 9–12 taking the 2023 AP CSP exam
Lone Star school	0.0310 (0.0557) 0.579	-0.0007 (0.0032) 0.838	-0.0171† (0.0089) 0.058	-0.0222* (0.0095) 0.022
Baseline measure of the outcome	0.4882*** (0.1042) 0.000	0.5329*** (0.1015) 0.000	3.7752*** (0.7643) 0.000	1.2188*** (0.2465) 0.000
Charter school	0.0396 (0.0591) 0.505	0.0072† (0.0037) 0.053	0.0113 (0.0101) 0.265	0.0235* (0.0110) 0.035
Total number of students enrolled	0.0000† (0.0000) 0.099	-0.0000 (0.0000) 0.159	-0.0000 (0.0000) 0.799	-0.0000 (0.0000) 0.309
Locale type suburb	-0.0453 (0.0432) 0.297	-0.0003 (0.0026) 0.905	0.0058 (0.0067) 0.389	0.0070 (0.0074) 0.351
Locale type rural or town	-0.0388 (0.0530) 0.467	-0.0014 (0.0031) 0.651	0.0167* (0.0080) 0.041	0.0198* (0.0089) 0.029
Proportion of students who are female	-0.1055 (0.2561) 0.681	-0.0049 (0.0150) 0.747	-0.0218 (0.0409) 0.595	-0.0094 (0.0440) 0.832
Proportion of students who are Black	-0.2063 (0.2954) 0.487	-0.0055 (0.0170) 0.747	-0.0616 (0.0456) 0.180	-0.0641 (0.0495) 0.199
Proportion of students who are Hispanic	-0.4370* (0.1982) 0.030	-0.0178 (0.0113) 0.119	-0.0240 (0.0312) 0.443	-0.0166 (0.0329) 0.616
Proportion of students who are economically disadvantaged	-0.2386 (0.1886) 0.209	0.0030 (0.0110) 0.786	0.0745* (0.0299) 0.015	0.0673* (0.0322) 0.040
Missing proportion of students who are economically disadvantaged	0.1107 (0.0971) 0.258	0.0072 (0.0057) 0.205	-0.0207 (0.0152) 0.176	-0.0133 (0.0165) 0.420
Constant	0.6000*** (0.1743) 0.001	0.0210* (0.0088) 0.019	-0.0134 (0.0228) 0.558	-0.0041 (0.0253) 0.872
Observations	96	96	96	96
R-squared	0.633	0.596	0.590	0.627

***p < .001, **p < .01, *p < .05, †p < .10

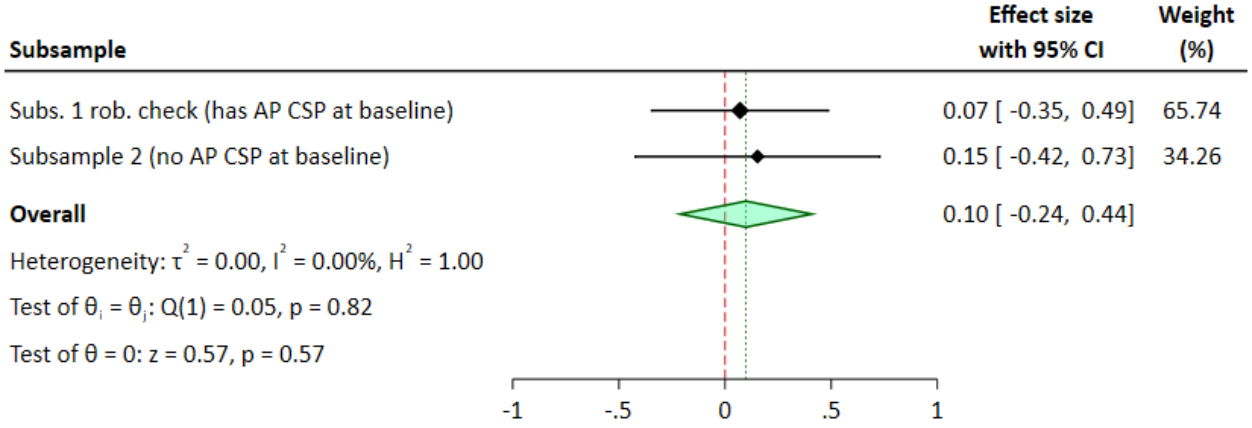
Note. Each cell contains the coefficient, the standard error (in parentheses), and the p value.

Exhibit G-10. Forest Plots from the Year 1 Meta-Analysis with Robustness Check Subsample 1

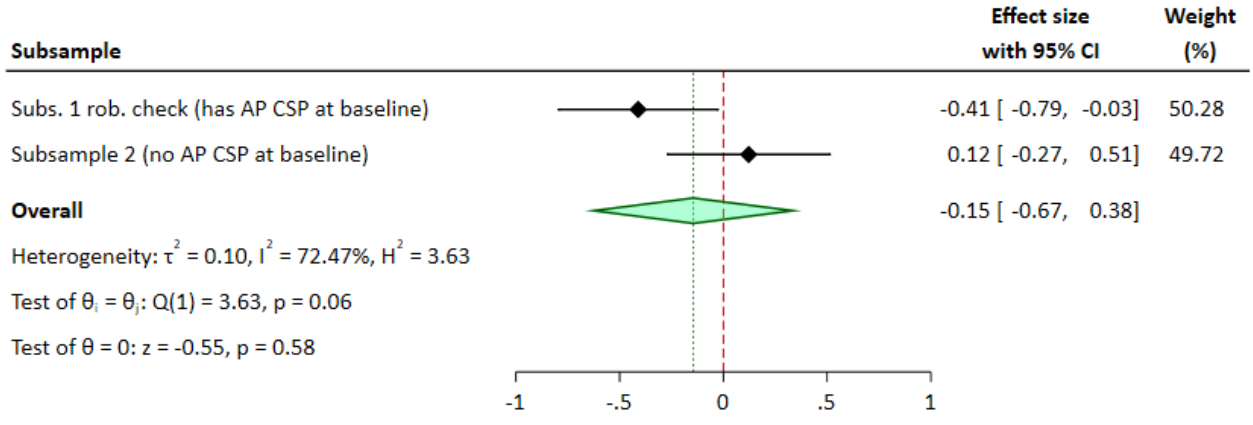
a. Percentage of 2021 AP CSP exam takers earning a 3, 4, or 5 on the exam



b. Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2021 AP CSP exam



c. Percentage of students in grades 9–12 earning a 1 or 2 on the 2021 AP CSP exam



d. Percentage of students in grades 9–12 taking the 2021 AP CSP exam

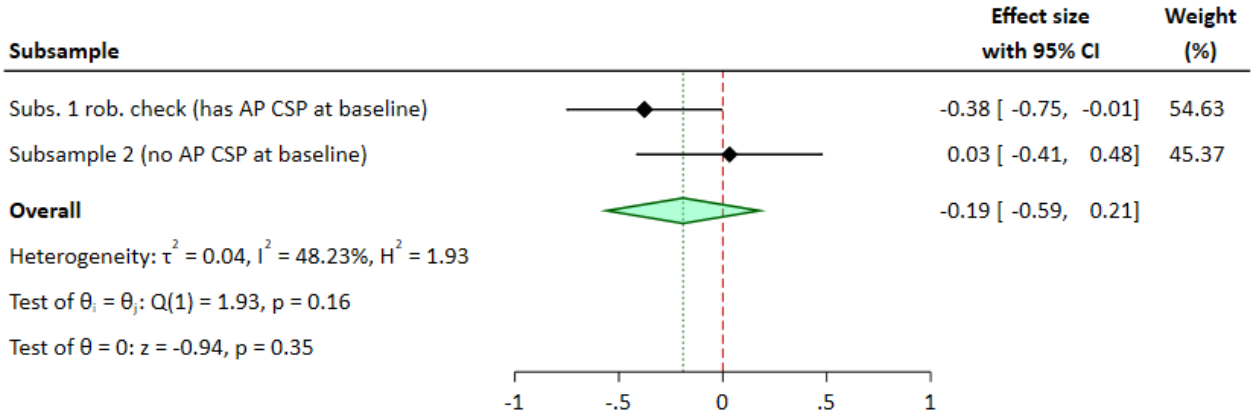
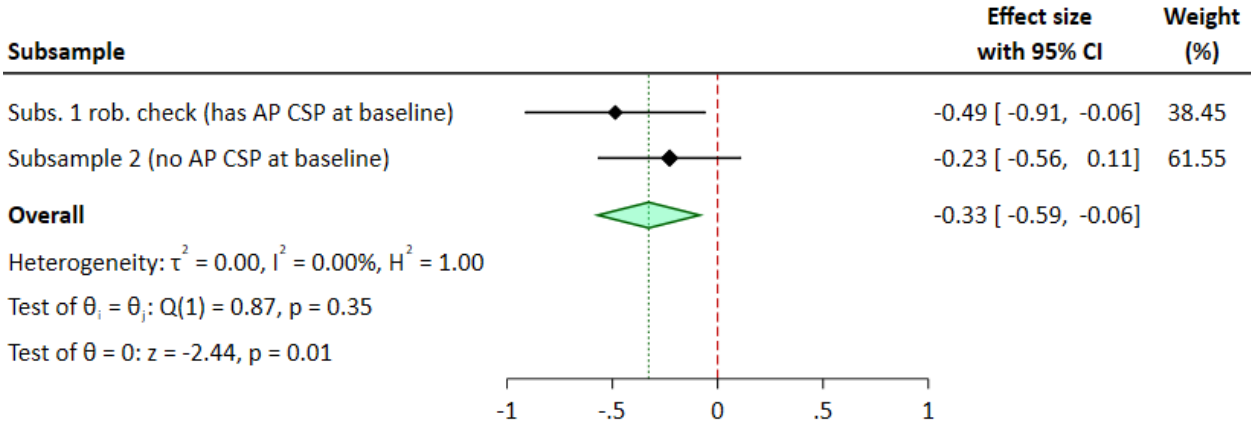


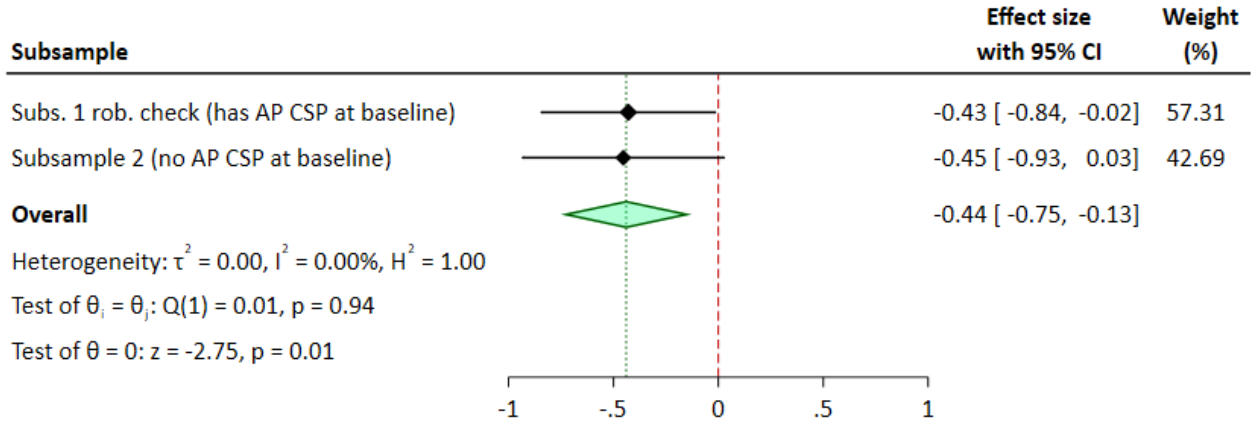
Exhibit G-11. Forest Plots from the Year 2 Meta-Analysis with Robustness Check Subsample 1

a. Percentage of 2022 AP CSP exam takers earning a 3, 4, or 5 on the exam



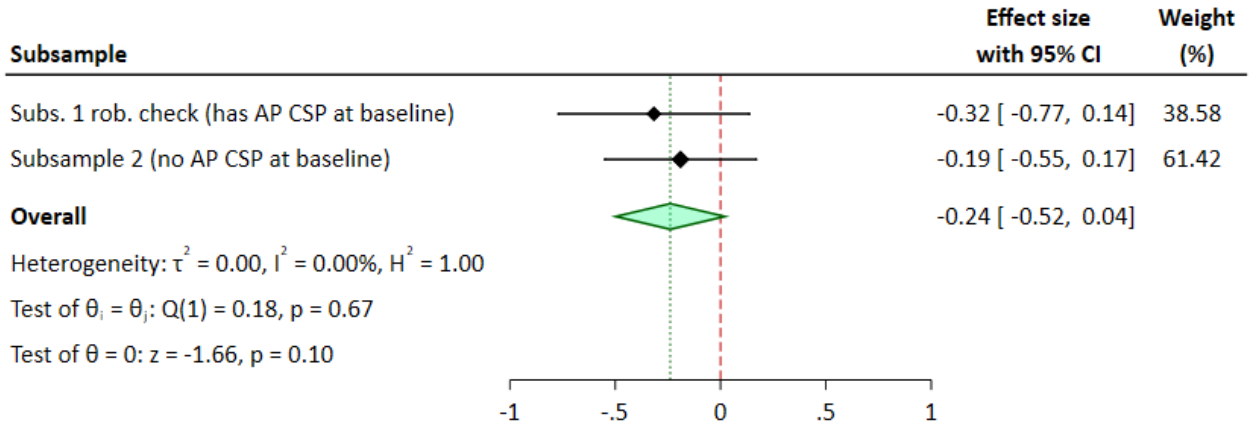
Note. For test of θ, $p = 0.0146$.

b. Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2022 AP CSP exam



Note. For test of θ, $p = 0.0059$.

c. Percentage of students in grades 9–12 earning a 1 or 2 on the 2022 AP CSP exam



Note. For test of θ, $p = 0.0963$.

d. Percentage of students in grades 9–12 taking the 2022 AP CSP exam

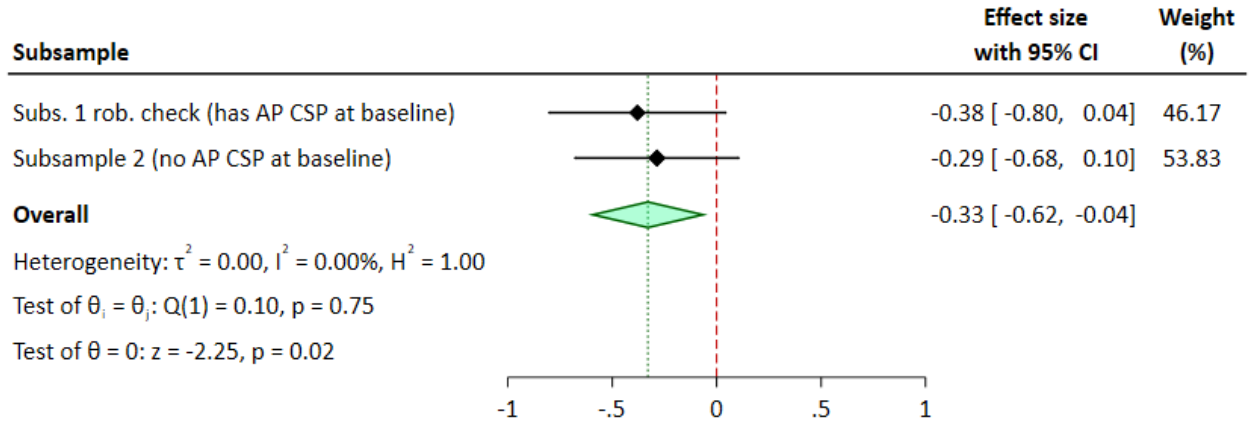
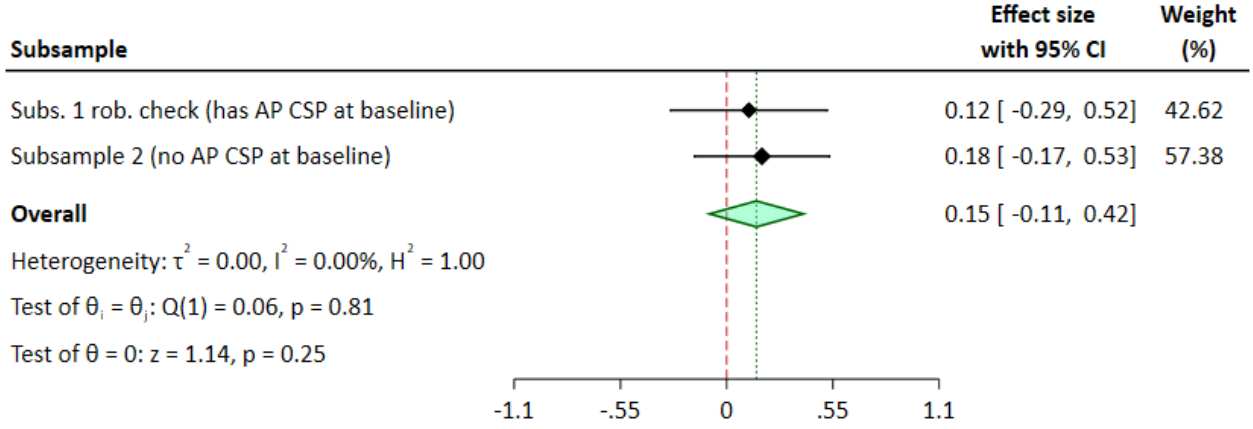
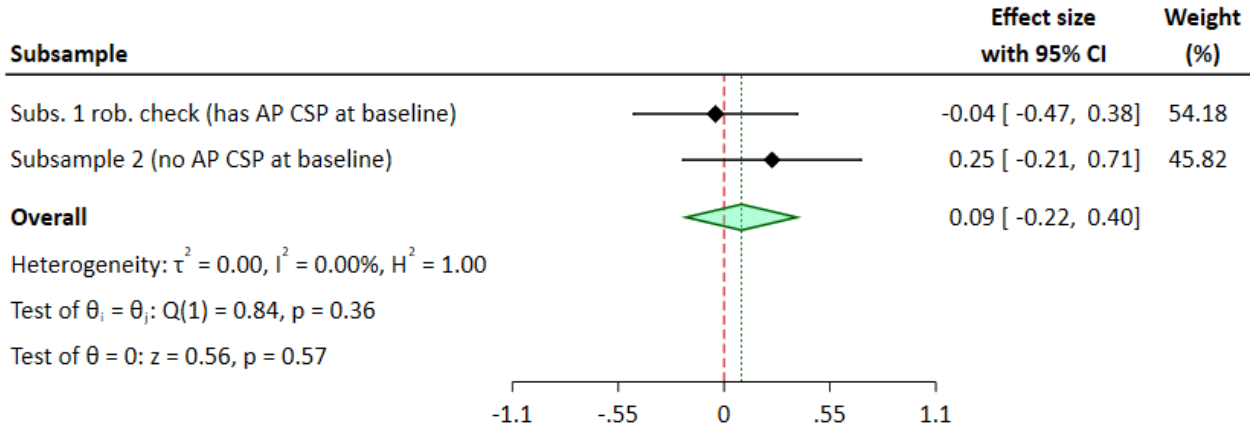


Exhibit G-12. Forest Plots from the Year 3 Meta-Analysis with Robustness Check Subsample 1

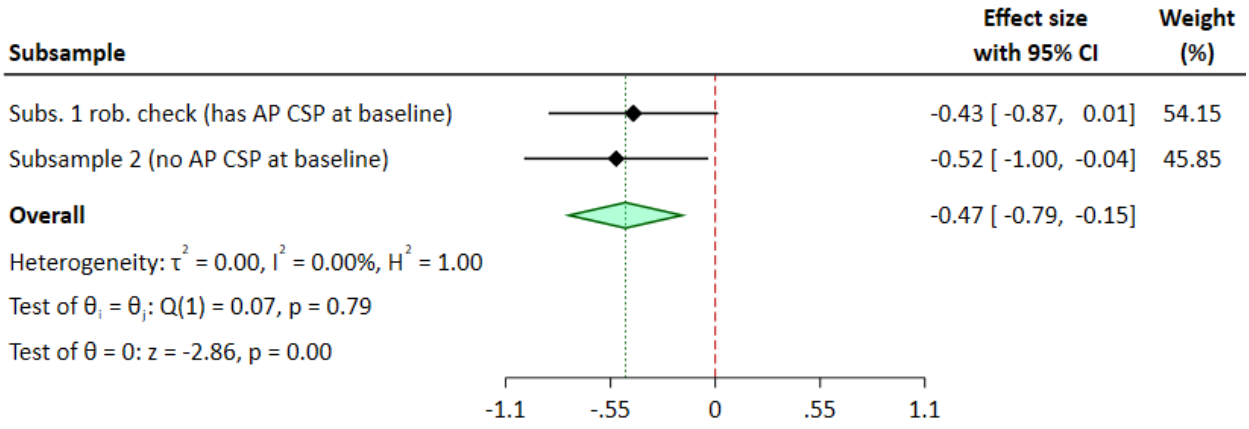
a. Percentage of 2023 AP CSP exam takers earning a 3, 4, or 5 on the exam



b. Percentage of students in grades 9–12 earning a 3, 4, or 5 on the 2023 AP CSP exam

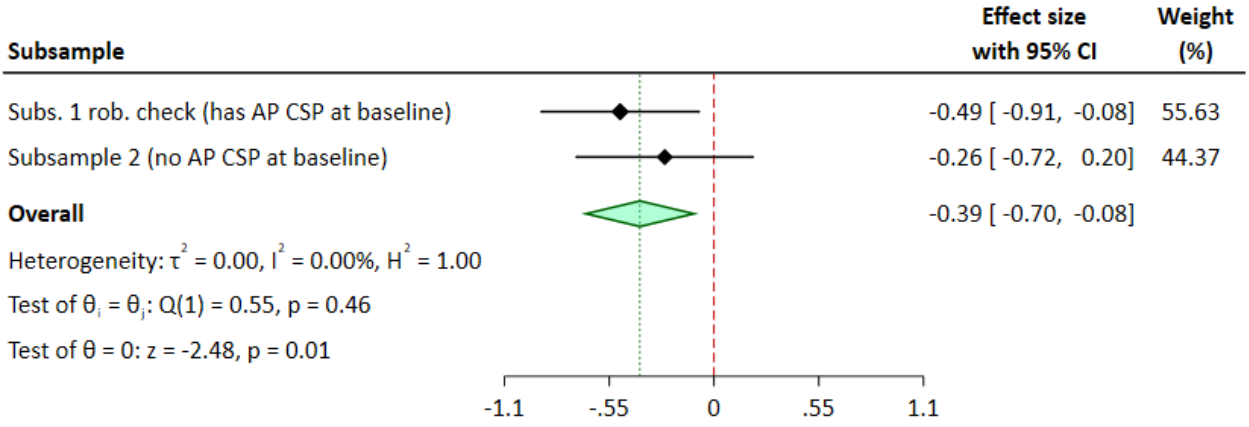


c. Percentage of students in grades 9–12 earning a 1 or 2 on the 2023 AP CSP exam



Note. For test of θ, $p = 0.0042$.

d. Percentage of students in grades 9–12 taking the 2023 AP CSP exam



Note. For test of θ, $p = 0.0132$.

About the American Institutes for Research®

Established in 1946, the American Institutes for Research® (AIR®) is a nonpartisan, not-for-profit institution that conducts behavioral and social science research and delivers technical assistance both domestically and internationally in the areas of education, health, and the workforce. AIR's work is driven by its mission to generate and use rigorous evidence that contributes to a better, more equitable world. With headquarters in Arlington, Virginia, AIR has offices across the U.S. and abroad. For more information, visit AIR.ORG.

AIR® Headquarters
1400 Crystal Drive, 10th Floor
Arlington, VA 22202-3289
+1.202.403.5000 | **AIR.ORG**

28467_08/25