Girls Who Code Program Evaluation

Final Report

Shuqiong (Linda) Lin, Megha Joshi, Kate Caton, Joseph P. Wilson
American Institutes for Research

March 2024

AIR®

Advancing Evidence.
Improving Lives.

Contents

Exhibits

Executive Summary

The American Institutes for Research® (AIR®) partnered with Girls Who Code (GWC) to conduct an independent study evaluating the effectiveness of two GWC summer virtual programs: the Summer Immersion Program (SIP) and the Self-Paced Program (SPP). These two programs are designed to enhance high school female and nonbinary students' interest and proficiency in computer science (CS) and to encourage the pursuit of CS-related higher education and careers.

The study used a quasi-experimental design to answer two research questions: (1) What are the effects of participation in GWC's SIP and SPP on majoring in a CS-related field during postsecondary education? and (2) How do these program effects vary by different student groups? Using publicly available National Student Clearinghouse data and GWC program records for the years 2020–22, the study compared SIP and SPP participants to similar waitlisted students to determine the effects of program participation on majoring in a CS-related field. The study yielded two key findings.

- First, on average, both SIP and SPP participants were significantly more likely to major in a CS-related field (by 13.2 percentage points and 11.5 percentage points, respectively) than comparison students.
- Second, both SIP and SPP consistently demonstrated positive effects on majoring in a CS-related field across most of the student groups examined, including White, Black or African American, and Hispanic or Latinx students; students who are historically underrepresented in computing; and students with little to no prior CS knowledge.

To optimize program outcomes, the study team suggests that GWC to (a) explore supplementary feedback mechanisms to gather insights from participants, including program exit interviews, focus groups and long-term alumni surveys; and (b) establish a comprehensive evaluation system to track the program's impact on various outcomes by including additional mid-term outcomes, such as enrollment in computer science (CS)-related Advanced Placement (AP) courses and successful completion of these courses.

This study comes with several limitations that warrant consideration when interpreting findings. First, estimated program effects may be influenced by a few preexisting student characteristics (e.g., GPA and motivation) that we were not able to access and control for in the analyses. Second, certain student groups had small sample sizes, resulting in low statistical power for analyzing variations in program effects.

Introduction

Girls Who Code (GWC) is an international nonprofit organization that aims to close the gender gap in technology, inspire and educate students who are historically underrepresented in computer science (CS), and equip those students with the necessary computing skills to pursue CS-related education and workforce opportunities. To achieve this, GWC provides two free virtual summer programs—the Summer Immersion Program (SIP) and the Self-Paced Program (SPP)—to prepare high school students who identify as girls or nonbinary for CS-related postsecondary education and technology careers. The American Institutes for Research® (AIR®) partnered with GWC to conduct an independent, quasi-experimental study to investigate program effects on the pursuit of CS-related postsecondary education. This report provides a brief introduction to the study background, an overview of the GWC summer programs, details on the study design, and a summary and discussion of key findings.

Study Background

Computing and information technology jobs are projected to grow by 15% between 2021 and 2031, making computing one of the fastest growing occupations in the U.S. labor market (Bureau of Labor Statistics, 2022). Women remain underrepresented in science, technology, engineering, and mathematics (STEM) fields, accounting for just 27% of today's STEM workforce (Martinez & Christnacht, 2021). Although women's participation in STEM fields has increased over the last 30 years, participation in CS specifically has declined from 37% in 1995 to 22% in 2022 (Girls Who Code, 2022; Zweben & Bizot, 2022). Given these labor market projections and the current composition of CS occupations, it is essential to invest in diversifying the U.S. workforce so that the general U.S. population is represented in future CS jobs. Research has shown that women's limited participation in CS is driven by gender stereotypes; male-dominated cultures; isolation; and lack of interest, confidence, and role models (Bejerano & Bartosh, 2015; Cheryan et al., 2020; Dasgupta & Stout, 2014; Google Inc. & Gallup Inc., 2016; Kröhn et al., 2020). It is therefore crucial to change women's perception of CS and provide trainings that can pique their interest in learning CS and prepare them with the necessary skills to pursue CS-related postsecondary education and careers.

GWC Summer Programs

In an effort to address gender disparities in the CS field and promote diversity in the CS pipeline, GWC offers two free virtual summer programs, SIP and SPP, to high school students, specifically designed for female and nonbinary students. These GWC summer programs aim to elevate students' proficiency in and enthusiasm for CS and encourage the pursuit of CS-related higher education and careers. Program participants have the opportunity to (a) learn coding through hands-on, real-world projects in game design, cybersecurity, data science, web development, and artificial intelligence (AI); (b) build sisterhood and make meaningful friendships through a lifelong alumni network; and (c) gain exposure to technology careers by engaging with engineers and entrepreneurs and participating in skill-building workshops.

GWC SUMMER PROGRAMS

SIP. GWC introduced the virtual SIP in 2020. This 2-week synchronous live program was created in response to the COVID-19 pandemic and serves as a replacement for the in-person SIP originally offered by GWC. Throughout the program, participants are introduced to key CS concepts and guided through step-by-step tutorials to complete real-world projects. Participants also have the chance to engage in a mentorship workshop and gain insights into career opportunities in STEM from women actively working in the technology industry.

SPP. GWC launched SPP in 2021. This 6-week asynchronous virtual program offers the flexibility of self-directed online learning for students who want to learn at their own pace and prefer not to adhere to a set schedule. Through this program, students can choose among courses in cybersecurity, data science and AI, and web development to earn certificates in their chosen CS tracks. Students also have the opportunity to build community through weekly live advisory sessions and GWC partner events.

Evaluation Design

Research Questions

AIR's evaluation was designed to investigate the effects of participating in GWC programs on majoring in a CS-related field during postsecondary education. The evaluation addresses the following research questions (RQs):

RQ1. What are the effects of participation in GWC's SIP and SPP on majoring in a CS-related field during postsecondary education?

RQ2. How do program effects vary by student background characteristics (i.e., historically underrepresented group status, student race/ethnicity, eligibility for free or reduced-price lunch, and prior CS knowledge level)?

Data Sources

This study used data that GWC collected from the following two sources:

- **National Student Clearinghouse data.** These data contained postsecondary enrollment information for program participants and for waitlist students who applied but were not admitted during the summers of 2020, 2021, and 2022 (i.e., cohorts 2020 to 2022). GWC used this information to determine students' postsecondary majors and whether those majors were related to CS or not. [1]
- **GWC internal program records.** These data included information on program type, student cohort, and student background characteristics (such as eligibility for free or reduced-price lunch and prior CS knowledge level) for program participants and waitlist students in cohorts 2020 to 2022.[2] GWC collected these data directly from program participants and waitlist students via GWC program application forms.

Quasi-Experimental Design

The AIR team used a quasi-experimental design with inverse propensity score weighting to answer the study's research questions. Below, we describe the analytic samples used in the study, the technique used to control for preexisting differences between program participants and waitlist students, and the method used to estimate program effects.

Analytic Sample. The treatment group comprised SIP participants from the 2020 to 2022 cohorts and SPP participants from the 2021 and 2022 cohorts. Students from the same cohorts who applied for the same program but were waitlisted served as the comparison groups.[3] In total, the analytic sample for SIP comprised 2,685 treatment students and 2,786 comparison students. The analytic sample for SPP comprised 1,027 treatment students and 797 comparison students.[4] Exhibit 1 shows the analytic sample sizes and background characteristics of treatment and comparison groups in this study. For SIP, the treatment group had a higher percentage of students who were eligible for free or reduced-price lunch, students who were members of historically underrepresented groups, and participants who enrolled in the summer after their junior year. The comparison group had a higher percentage of Asian students and students who applied in the summer after their sophomore year. For SPP, the

---

[1] GWC's major classification framework is based on both CIP codes and the name of the major provided by the college or university. GWC validates major classifications annually, maintaining a >95% accuracy rate.

[2] Program type refers to either the program participants applied to and enrolled in (SIP or SPP) or the program to which waitlist students applied. GWC collected and generated all variables from this data source.

[3] We conducted similar analyses for other GWC summer programs: a 1-week SIP, a pilot model implemented only in 2020; and an in-person SIP for the 2019 cohort. Please see Appendix D for findings on these other GWC summer programs.

[4] The analytic sample was limited to U.S. students. Six students had their states listed as one of the American territories and 14 students had no information regarding their state. These students were removed from all analytic samples.

treatment group had a higher percentage of Asian students. The comparison group had a higher percentage of Hispanic or Latinx students and students with no prior knowledge of CS.

Exhibit 1. SIP and SPP Treatment and Comparison Samples

	SIP		SPP	
	Treatment (*N* = 2,685)	Comparison (*N* = 2,786)	Treatment (*N* = 1,027)	Comparison (*N* = 797)
Cohort				
Summer 2020 sample size	612	1,009	n/a	n/a
Summer 2021 sample size	2,062	1,769	571	602
Summer 2022 sample size	11	8	456	195
Eligible for free or reduced-price lunch				
Yes	49.2%	38.3%	39.6%	38.6%
Not eligible or unsure	50.8%	61.7%	60.4%	61.4%
Race/ethnicity				
Asian	31.1%	43.2%	51.2%	44.5%
Black/African American	20.0%	16.1%	15.8%	15.6%
Hispanic/Latinx	20.8%	18.5%	13.3%	17.8%
Multiracial	5.7%	4.1%	4.2%	4.0%
White	19.4%	15.2%	12.4%	15.6%
Additional (small participant sizes)	3.1%	2.9%	2.7%	1.9%
Historically underrepresented students	69.7%	53.3%	54.0%	54.6%
Age – mean (standard deviation)	19.0 (0.7)	18.9 (0.7)	18.6 (0.6)	18.6 (0.6)
Prior CS knowledge				
No prior knowledge	74.0%	70.3%	28.2%	33.9%
Beginner	7.2%	4.6%	22.7%	16.2%
Intermediate	10.3%	13.6%	19.1%	15.2%
Advanced	8.5%	11.6%	30.0%	34.8%
Grade				
Summer after sophomore year	22.8%	36.2%	0%	0%
Summer after junior year	76.8%	63.5%	55.6%	75.5%
Summer after senior year	0.5%	0.3%	44.4%	24.5%

Note. n/a = not applicable. Students whose free or reduced-price lunch status was “Unsure” were combined with students who were ineligible for free or reduced-price lunch during the analyses, given the small sample size for the “Unsure” category. The “Additional” race and ethnicity subgroup included Native Hawaiian or Pacific Islander, American Indian or Alaska Native, and any other race categories that lacked a sufficient number of observations (less than 3% of the sample). Historically underrepresented students were defined as those whose race was not White or Asian, who qualified for free or reduced-price lunch, who received a SIP stipend, or who were first-generation college students.

Inverse Propensity Score Weighting. To control for differences between treatment and comparison groups, as shown in Exhibit 1, we used inverse propensity score weighting. Propensity scores were estimated using generalized Bayesian additive regression trees (BART), an advanced machine learning technique that has been shown to provide more accurate estimation than traditional statistical models (Hill et al., 2020). With inverse propensity score weighting, we achieved baseline equivalence on all key baseline variables that may have influenced both students' participation in the GWC program and the outcome, including students' eligibility for free or reduced-price lunch, race/ethnicity, prior CS knowledge level, and cohorts (see Appendix A for the list of baseline covariates used in this study). We also conducted sensitivity analyses to examine the robustness of our findings to the choice of propensity score estimation method. For a detailed discussion of the data cleaning process, analytic methods, and any other technical details, including sensitivity checking results, please see Appendix B.

Program Effectiveness Estimation. We estimated program effects on outcomes for each of the GWC programs (RQ1) using a weighted least squares linear probability model with clustered standard errors (students clustered in states). The model contained all baseline characteristics that were used to estimate propensity scores to account for any remaining differences in observed characteristics between treatment and comparison students. To understand how program effects varied by students' preexisting characteristics (i.e., historically underrepresented group status, student race/ethnicity, eligibility for free or reduced-price lunch, and prior CS knowledge level; RQ2), we calculated program effects for each student group and conducted post-hoc pairwise contrast tests to examine whether program effects significantly differed for one student group versus another.

Evaluation Results

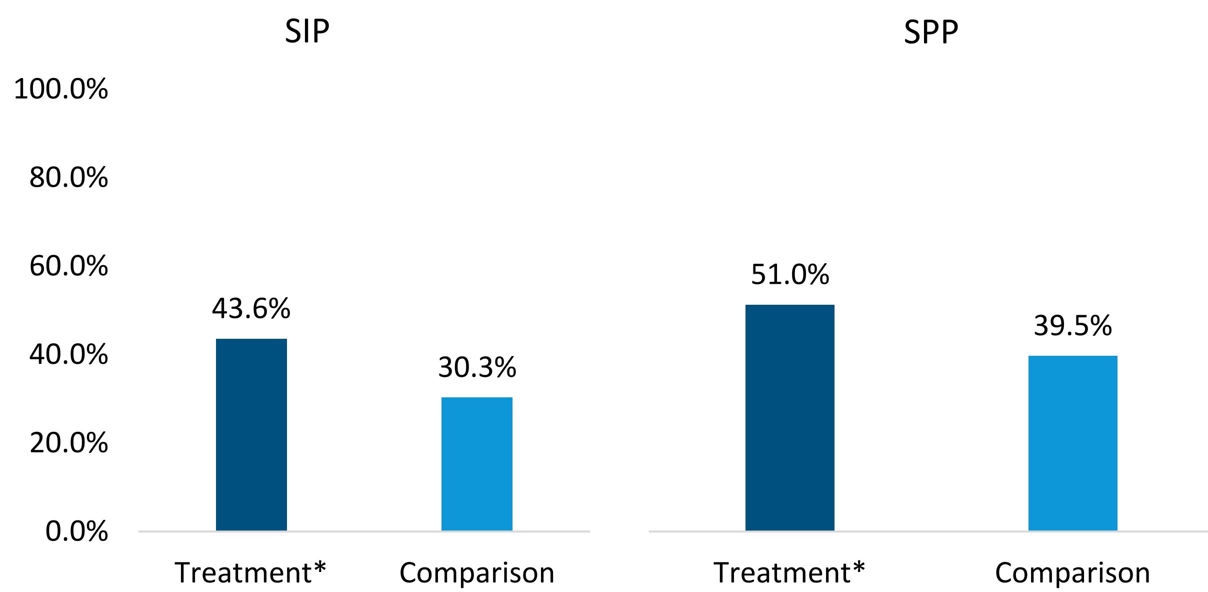
Overall Program Effects

Overall, students who participated in SIP or SPP were significantly more likely to major in a CS-related field than comparison students (Exhibit 2; Appendix C). Specifically, on average, students who participated in SIP were significantly more likely to major in a CS-related field (by 13.2 percentage points) than SIP waitlist students. The adjusted percentage of students who majored in a CS-related field was 43.6% for SIP participants and 30.3% for the comparison group.[5] Students who participated in SPP were also significantly more likely to major in a CS-related field (by 11.5 percentage points) than comparison students. The adjusted percentage of

[5] The adjusted percentage is the estimated percentage of students majoring in a CS-related field after we implemented inverse propensity score weighting and regression adjustment for all the covariates.

students who majored in a CS-related field was 51.0% for SPP participants and 39.5% for the comparison group.

Exhibit 2. Adjusted Percentage of Students Majoring in a CS-Related Field



* The program effect is statistically significant at α = .05.

Program Effects by Student Group

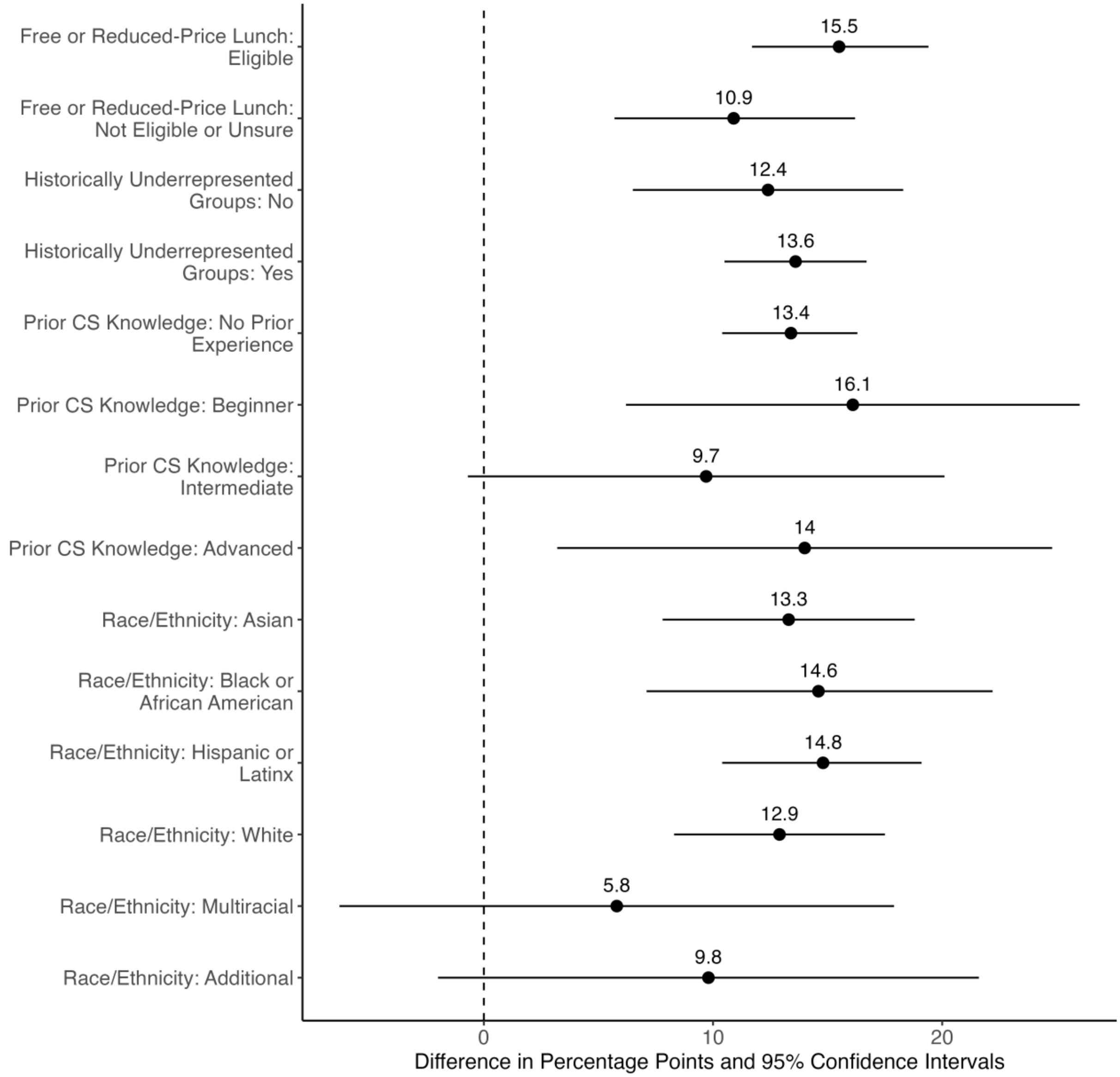
Positive program effects on majoring in a CS-related field were evident not only across the entire student sample, but also within most of the student groups. SIP participants were significantly more likely to start majoring in a CS-related field compared to comparison students, regardless of students' race/ethnicity, historical underrepresentation in computing, and eligibility for free or reduced-price lunch.[6] When looking at students with different levels of prior CS knowledge, SIP had significantly positive effects for those who had no prior CS knowledge or were beginners. For those with intermediate or advanced prior knowledge, the effects were positive but not statistically significant. Similar conclusions were drawn for SPP, with the exception of participants who were eligible for free or reduced-price lunch and Asian students, who demonstrated positive but non-significant program effects. Exhibits 3 and 4 display SIP and SPP effects for each student group. Appendix C provides more detail on estimated program effects for each student group.

Additionally, the results of pairwise comparison analysis reveal variability in the magnitude of program effects across student groups, yet no statistically significant differences were detected.

[6] This conclusion does not incorporate the "Multiracial" and "Additional" race/ethnicity groups because of their extremely small sample sizes.

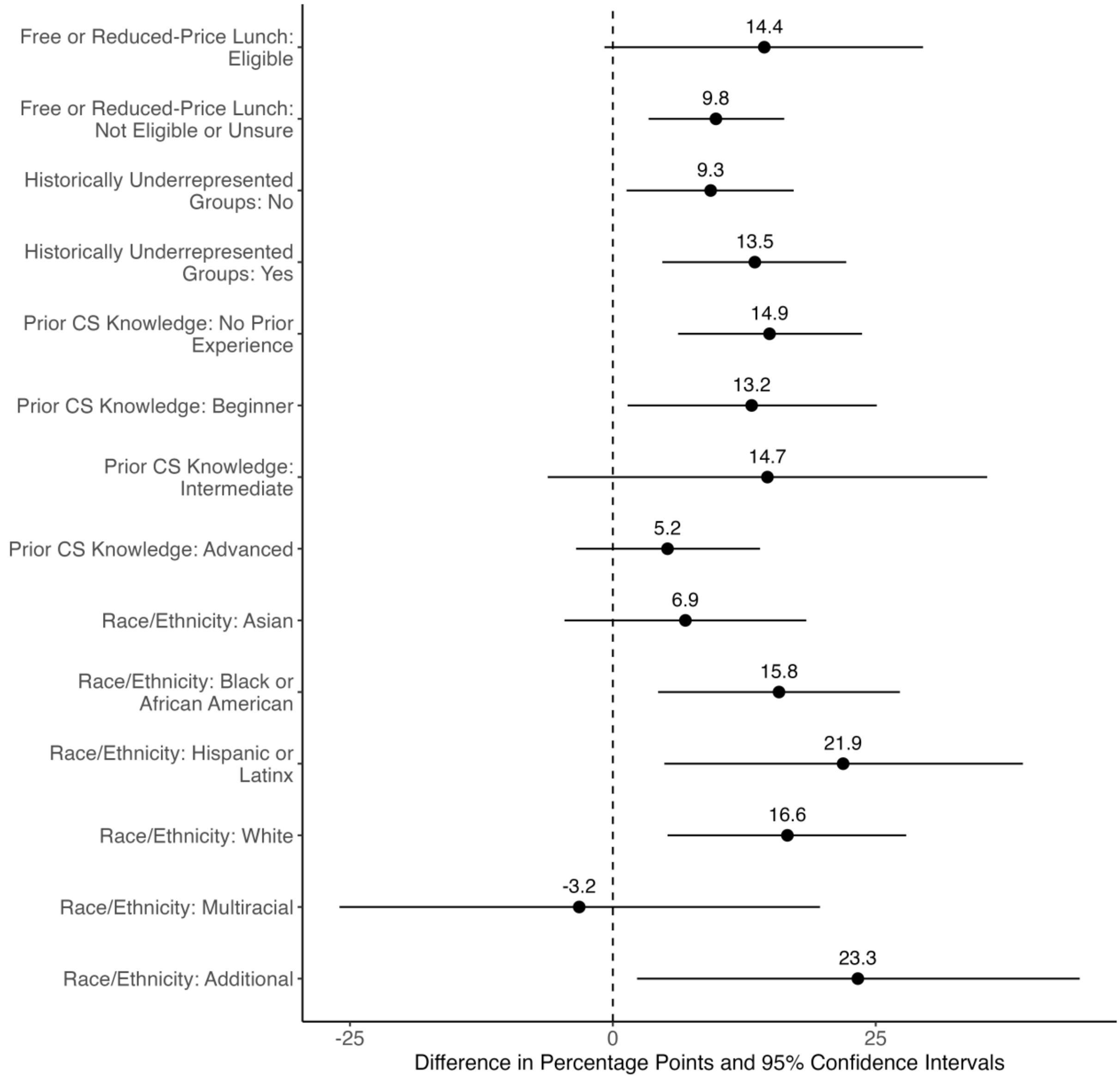
It is important to highlight that observed non-significant differences may be attributed to small sample sizes. The differences in program effects' magnitude may still have practical significance.

Exhibit 3. SIP Effects by Student Group



Note. The dots indicate the estimated program effect and the lines represent the 95% confidence intervals for estimated program effects.

Exhibit 4. SPP Effects by Student Group



Note. The dots indicate the estimated program effect and the lines represent the 95% confidence intervals for estimated program effects.

Conclusions and Considerations

To evaluate the effects of participating in GWC's SIP and SPP on students majoring in a CS-related field, this study compared SIP and SPP participants to students who were waitlisted but were otherwise similar to program participants. We found that SIP and SPP participants were more likely to major in a CS-related field, both overall and across most student groups. Based on the study's findings, we would like to propose the following program improvement suggestions to further expand the impact of SIP and SPP, and create more effective educational experience for those students who were historically underrepresented in the CS field.

Explore additional mechanisms to collect feedback from participants to further refine the programs. Currently, GWC conducts post-program surveys to evaluate participants and teaching teams ' overall satisfaction, satisfaction with various curriculum components and program activities/resources, challenges encountered during the program, and suggestions for enhancement. To enhance the depth of feedback, we propose the integration of additional feedback mechanisms alongside the existing infrastructure. These mechanisms aim to gather direct insights from all participants regarding their learning experiences, particularly focusing on identifying successes and overcoming barriers encountered by student groups who experienced relatively larger and smaller program effects. We recommend implementing the following supplementary feedback mechanisms:

- Program exit interviews and focus groups: Tailored for specific student groups, these sessions allow for in-depth discussions on participants' experiences, challenges, and suggestions for improvement.
- Long-term alumni surveys: Conducted after a significant period since program completion, these surveys gather insights into alumni's post-program experiences, career trajectories, and the program's influence on their educational or professional paths.

Establishment of a comprehensive monitoring and evaluation system to continuously track the impact of the program on diverse outcomes. While current program evaluation primarily centers on short and long-term effects, including students' perceived interest and proficiency in computer science (CS) and attainment of CS-related college degrees, we advocate for additional attention to mid-term outcomes. These mid-term outcomes encompass program participants' behaviors and performance related to high school CS education, such as pursuit of secondary education pathway in CS-related fields, enrollment in CS-related Advanced Placement (AP) courses, successful completion of these courses, and participation in CS competitions. This expanded focus will provide a more holistic understanding of the program's effectiveness and enable targeted interventions to enhance outcomes across various stages of program participants' CS-related education journeys.

Limitations and Future Exploration

We acknowledge that this study experienced some constraints and note that the findings presented in this report should be interpreted with some caution.

- **Limited confounding variables.** We accounted for numerous key confounding variables for which GWC had data. However, certain additional cognitive and noncognitive confounding variables—including high school GPA, motivation, and self-efficacy—were not accessible or accounted for in the analysis. These variables could potentially bias the study's estimated effects if they influence the likelihood of students applying to a GWC program, being selected into the program, or choosing a major in a CS-related field. As a result, we urge caution in interpreting the findings as causal impacts of the GWC programs. *To enhance the study's rigor, we recommend collecting data and accounting for a more comprehensive list of confounding variables, or conducting a randomized control trial where students are randomly selected to participate in the programs.*
- **Insufficient statistical power to detect treatment effect variation.** For certain student group analyses, such as the "Multiracial" and "Additional" race/ethnicity groups, and analyses detecting program effects difference among student groups, the sample size was considerably smaller, possibly resulting in low power to detect potential program effects or potential program effect differences among groups. *To better understand program effects on these student groups and program effect differences, we recommend aggregating data from additional cohorts to increase the analytical power of the analysis.*

References

Bejerano, A. R., & Bartosh, T. M. (2015). Learning masculinity: Unmasking the hidden curriculum in science, technology, engineering, and mathematics courses. *Journal of Women and Minorities in Science and Engineering*, *21*, 107–124.

Bureau of Labor Statistics. (2022, September 8). *Computer and information technology occupations*. Office of Occupational Statistics and Employment Projections. https://www.bls.gov/ooh/computer-and-information-technology/home.htm#:~:text=Overall%20employment%20in%20computer%20and,new%20jobs%20over%20the%20decade

Cheryan, S., Lombard, E. J., Hudson, L., Louis, K., Plaut, V. C., & Murphy, M. C. (2020). Double isolation: Identity expression threat predicts greater gender disparities in computer science. *Self and Identity*, *19*, 412–434.

Dasgupta, N., & Stout, J. (2014). Girls and women in science, technology, engineering, and mathematics. *Policy Insights From the Behavioral and Brain Sciences*, *1*(1), 21–29. https://doi.org/10.1177/2372732214549471

Girls Who Code. (2022). *The gender gap in computing is getting worse*. https://girlswhocode.com/about-us

Google Inc. & Gallup Inc. (2016). *Diversity gaps in computer science: Exploring the underrepresentation of girls, Blacks and Hispanics*. http://goo.gl/PG34aH

Hill, J., Linero, A., & Murray, J. (2020). Bayesian additive regression trees: A review and look forward. *Annual Review of Statistics and Its Application*, *7*, 251–278.

Kröhn, C., Groher, I., Sabitzer, B., & Kuka, L. (2020). Female computer scientists needed: Approaches for closing the gender gap. *2020 IEEE Frontiers in Education Conference* (FIE), 1–4.

Martinez, A., & Christnacht, C. (2021, January 26). *Women are nearly half of U.S. workforce but only 27% of STEM workers. Women making gains in STEM occupations but still underrepresented*. U.S. Census Bureau. https://www.census.gov/library/stories/2021/01/women-making-gains-in-stem-occupations-but-still-underrepresented.html

Zweben, S., & Bizot, B. (2022, May). *2021 Taulbee Survey: CS enrollment grows at all degree levels, with increased gender diversity*. Computing Research Association. https://cra.org/wp-content/uploads/2022/05/2021-Taulbee-Survey.pdf

Appendix A. Baseline Equivalence

The study team assessed baseline balance before and after applying inverse propensity score weighting to a variety of key student-level baseline characteristics that were available, including program cohort, students' eligibility for free or reduced-price lunch (FRPL), race/ethnicity, an indicator for whether a student was part of historically underrepresented groups (HUGS) in CS, age, prior CS knowledge level, and students' grade. We used a standardized mean difference between -0.25 and 0.25 as the threshold for determining baseline equivalence.

For both programs, the differences in all baseline background characteristics between treatment and comparison groups were close to zero after weighting, indicating baseline equivalence between treatment and comparison groups. Exhibit A1 and Exhibit A2 depict the results of the balance assessment prior to and after re-weighting the comparison group on all baseline measures used in this study.

Exhibit A1. Estimated Standardized Mean Differences Between Treatment and Comparison Students on Baseline Measures for SIP

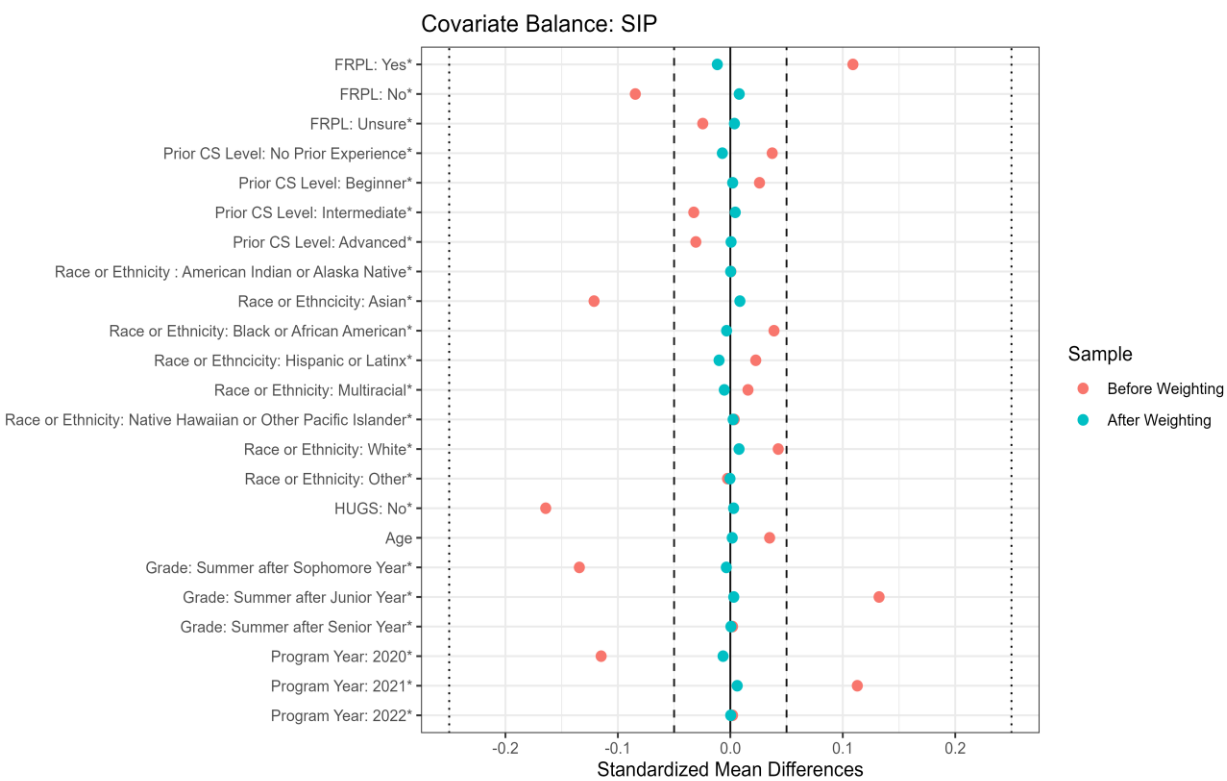
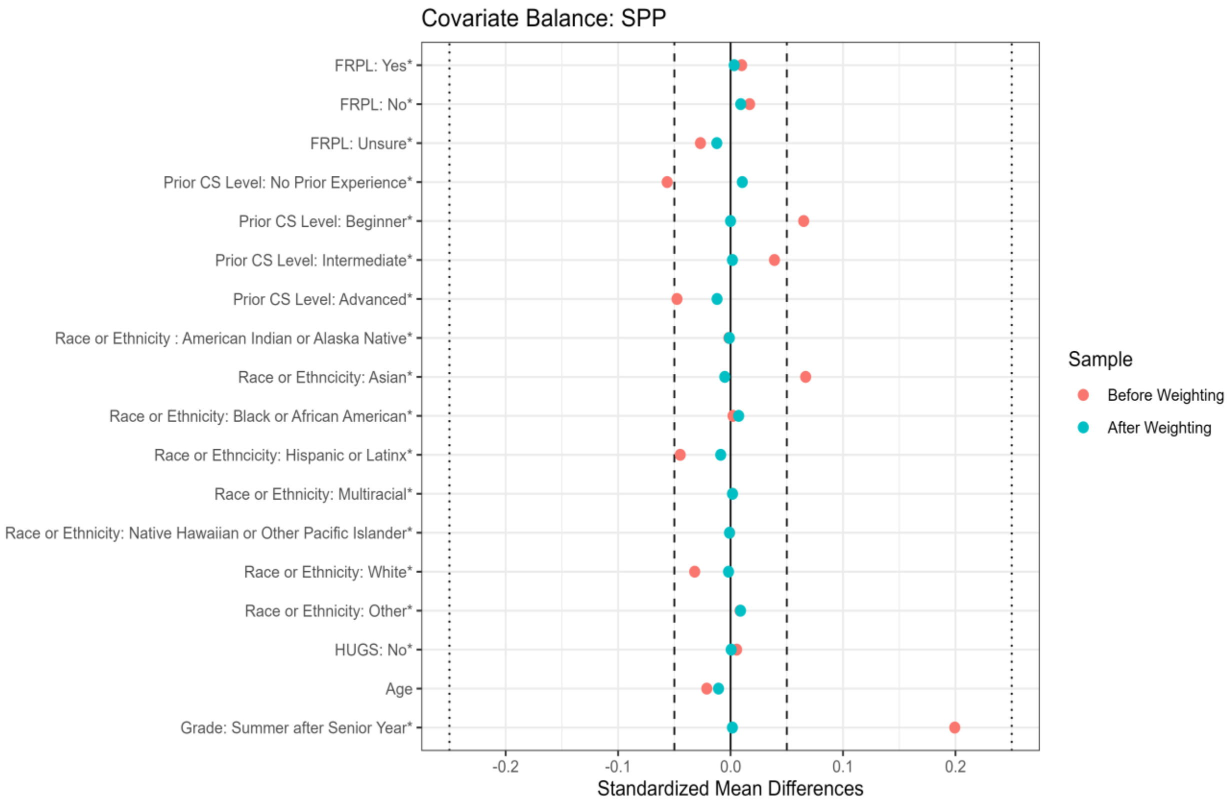


Exhibit A2. Estimated Standardized Mean Differences Between Treatment and Comparison Students on Baseline Measures for SPP



Appendix B. Technical Details

Data Preparation

Data were cleaned and analyzed using R programming. Input was received from GWC regarding grouping definitions for race, free or reduced-price lunch eligibility, and historically underrepresented groups. Duplicate observations were identified and dropped. Missing data were identified and either confirmed to be missing and excluded from the sample or supplemented with additional information provided by GWC. We conducted complete case analysis with any student with missing data in the outcome and/or any relevant background characteristics were excluded. While missing imputation for outcome variables is not generally recommended, low missing rates were found for most of the background characteristics, so using complete cases only was deemed reasonable. Several of the background characteristics, such as first-generation college student status, will be not used in the analyses due to high missing rates.

Propensity Score Estimation

We used an inverse propensity score weighting technique to create a balanced treatment and comparison group for estimating the causal impact of program participation on outcomes. To estimate propensity scores, we used Bayesian additive regression trees, a non-parametric model that does not impose strong assumptions about the functional form of the relationship between propensity scores and covariates (Hill et al., 2020). The propensity score model included all key student-level baseline characteristics that were available, including program cohort, students' eligibility for free or reduced-price lunch, race/ethnicity, an indicator for whether a student was part of historically underrepresented groups in CS, age, prior CS knowledge level, and students' grade. After estimating the propensity scores, we assessed the distributions of the estimated propensity scores for both treatment and comparison students to determine how comparable the groups were. The average treatment effect for the treated (ATT) weights were calculated as follows (Griefer, 2023):

$$w_{ij} = T_{ij} + (1 - T_{ij})\left(\frac{p_{ij}}{1 - p_i j}\right)$$

where w_{ij} denotes the ATT weight for student i in state j, T_{ij} is an indicator for whether the student is in the treatment group, and p_{ij} indicates the estimated propensity score for the student.

After estimating the weights, we checked for baseline equivalence by calculating weighted mean differences on all covariates to ensure that the comparison group of students was observationally equivalent to the treatment group of students (Griefer, 2023).

Impact Evaluation Model

To examine overall program effects (RQ1), we estimated the impact of the GWC programs on outcomes using a weighted least squares linear probability model with clustered standard errors. The model contained all of the covariates that were included in the propensity score model. The impact evaluation model was as follows:

$$Y_{ij} = \alpha + \beta T_{ij} + \gamma X_{ij} + \delta Cohort_{ij} + \epsilon_{ij}$$

Here Y_{ij} is the outcome for student i in state j; T_{ij} is an indicator for whether the student is in the treatment group; X_{ij} is a vector of student-level characteristics; $Cohort_{ij}$ denotes the program year to which the student belongs; and ϵ_{ij} is the student-level error term. We included all covariates in the outcome model to account for remaining variability in student characteristics. We conducted an impact evaluation analysis separately for each of the GWC programs.

To examine whether program effects differed across student groups (RQ2), we added an interaction term between the treatment indicator and the focal variable (i.e., historically underrepresented group status, student race/ethnicity, eligibility for free or reduced-price lunch, and prior CS knowledge level). We also conducted pairwise contrast tests to examine whether effects differed for one student group versus another.

Sensitivity Analyses

In addition to the propensity scores estimation method used in the main study, the study team applied a couple of different approaches for sensitivity checking purposes to understand the robustness of the findings. Specifically, we applied (1) traditional generalized linear modeling (GLM) to estimate the propensity scores, and (2) nearest nationhood matching with a 1:1 ratio (NN 1:1 matching) for sensitivity checking. We found that the program effects we estimated were similar across different methods, indicating the findings are robust (Exhibit B1).

Exhibit B1. Sensitivity Checking for SIP and SPP Program Effects

GWC program	Sensitivity checking method	Adjusted percentage of majoring in a CS-related field		Program effect	Standard error	*p*-value	Confidence interval
		Treatment	Comparison				
SIP	GLM inverse propensity score weighting	43.5	30.2	13.3	1.2	0	10.5–16
SIP	NN 1:1 matching	43.5	30.3	13.2	1.1	0	10.6–15.9
SPP	GLM inverse propensity score weighting	51.1	40.1	10.9	2.4	.001	5.6–16.2
SPP	NN 1:1 matching	49.3	39.3	10	2.6	.004	4.1–15.9

Appendix C. Program Effects for Each Student Group

Exhibit C1. Estimated SIP Effects by Student Group

Student groups	Adjusted percentage of majoring in a CS-related field		Program effect	Standard error	*p*-value	Confidence interval	
	Treatment	Comparison				Low	High
All students	43.6	30.3	13.2	1.2	<.001*	10.3	16.1
Historically underrepresented							
Yes	44	30.4	13.6	1.3	<.001*	10.5	16.7
No	45.5	33.1	12.4	2.7	.001*	6.5	18.3
Free/reduced-price lunch							
Eligible	43.2	27.7	15.5	1.5	<.001*	11.7	19.4
Not eligible or unsure	48.8	37.9	10.9	2.4	.001*	5.7	16.2
Race/ethnicity							
Asian	49.9	36.6	13.3	2.1	.002*	7.8	18.8
Black or African American	42	27.4	14.6	3.4	.002*	7.1	22.2
Hispanic or Latinx	41.9	27.2	14.8	1.8	<.001*	10.4	19.1
Multiracial	40	34.2	5.8	5.2	.294	-6.3	17.9
White	39.9	27	12.9	2.1	<.001*	8.3	17.5
Additional	44.4	34.7	9.8	4.8	.089	-2	21.6
Prior CS knowledge							
No prior knowledge	37.3	23.9	13.4	1.2	<.001*	10.4	16.3
Beginner	39.9	23.8	16.1	4.3	.005*	6.2	26
Intermediate	47.2	37.5	9.7	4.6	.063	-0.7	20.1
Advanced	55.4	41.4	14	4.9	.016*	3.2	24.8

* The program effect is statistically significant at $\alpha = .05$.

Exhibit C2. Estimated SPP Effects by Student Group

Student groups	Adjusted percentage of majoring in a CS-related field		Program effect	Standard error	*p*-value	Confidence interval	
	Treatment	Comparison				Low	High
All students	51.0	39.5	11.5	2.2	.001*	6.5	16.4
Historically underrepresented							
Yes	44.2	30.8	13.5	3.8	.008*	4.7	22.2
No	42.9	33.7	9.3	3.4	.027*	1.3	17.2
Free/reduced-price lunch							
Eligible	43.3	28.9	14.4	6.2	.059	-0.8	29.5
Not eligible or unsure	46.8	36.9	9.8	2.9	.007*	3.4	16.3
Race/ethnicity							
Asian	48.5	41.6	6.9	4.7	.193	-4.6	18.4
Black or African American	57.9	42.1	15.8	5.2	.012*	4.3	27.3
Hispanic or Latinx	48.2	26.3	21.9	7.1	.019*	4.9	39
Multiracial	44.7	47.9	-3.2	10.2	.762	-26	19.7
White	46.4	29.8	16.6	5.4	.006*	5.2	27.9
Additional	53.5	30.1	23.3	8.6	.035	2.3	44.4
Prior CS knowledge							
No prior knowledge	34.0	19.0	14.9	3.8	.004*	6.2	23.7
Beginner	39.8	26.6	13.2	5	.033*	1.4	25.1
Intermediate	45.2	30.5	14.7	9.2	.145	-6.2	35.6
Advanced	56.1	50.8	5.2	3.9	.213	-3.5	14

* The program effect is statistically significant at $\alpha = .05$.

Exhibit C3. Program Effect Differences for Pairwise Comparisons

Program	Variable	Pairwise comparison groups		Program effect for pairwise comparison groups		Program effect difference
SIP	**Historically underrepresented**	No	Yes	12.4	13.6	-1.2
	Race/ethnicity	Black or African American	Asian	14.6	13.3	1.3
		Hispanic or Latinx	Asian	14.8	13.3	1.5*
		Multiracial	Asian	5.8	13.3	-7.5
		White	Asian	12.9	13.3	-0.4*
		Additional	Asian	9.8	13.3	-3.5
		Hispanic or Latinx	Black or African American	14.8	14.6	0.2
		Multiracial	Black or African American	5.8	14.6	-8.8
		White	Black or African American	12.9	14.6	-1.7
		Additional	Black or African American	9.8	14.6	-4.8
		Multiracial	Hispanic or Latinx	5.8	14.8	-9
		White	Hispanic or Latinx	12.9	14.8	-1.9
		Additional	Hispanic or Latinx	9.8	14.8	-5
		White	Multiracial	12.9	5.8	7.1
		Additional	Multiracial	9.8	5.8	4
		Additional	White	9.8	12.9	-3.1
	Free/reduced-price lunch	Not eligible or unsure	Eligible	10.9	15.5	-4.6*
	Prior CS level	Beginner	No prior knowledge	16.1	13.4	2.7
		Intermediate	No prior knowledge	9.7	13.4	-3.7
		Advanced	No prior knowledge	14	13.4	0.6*
		Intermediate	Beginner	9.7	16.1	-6.4*
		Advanced	Beginner	14	16.1	-2.1*
		Advanced	Intermediate	14	9.7	4.3
SPP	**Historically underrepresented**	No	Yes	9.3	13.5	-4.2
	Race/ethnicity	Black or African American	Asian	15.8	6.9	8.9
		Hispanic or Latinx	Asian	21.9	6.9	15*
		Multiracial	Asian	-3.2	6.9	-10.1
		White	Asian	16.6	6.9	9.7

Program	Variable	Pairwise comparison groups		Program effect for pairwise comparison groups		Program effect difference
		Additional	Asian	23.3	6.9	16.4
		Hispanic or Latinx	Black or African American	21.9	15.8	6.1*
		Multiracial	Black or African American	-3.2	15.8	-19
		White	Black or African American	16.6	15.8	0.8
		Additional	Black or African American	23.3	15.8	7.5
		Multiracial	Hispanic or Latinx	-3.2	21.9	-25.1*
		White	Hispanic or Latinx	16.6	21.9	-5.3
		Additional	Hispanic or Latinx	23.3	21.9	1.4
		White	Multiracial	16.6	-3.2	19.8
		Additional	Multiracial	23.3	-3.2	26.5
		Additional	White	23.3	16.6	6.7
	Free/reduced-price lunch	Not eligible or unsure	Eligible	9.8	14.4	-4.6
	Prior CS level	Beginner	No prior knowledge	13.2	14.9	-1.7
		Intermediate	No prior knowledge	14.7	14.9	-0.2
		Advanced	No prior knowledge	5.2	14.9	-9.7*
		Intermediate	Beginner	14.7	13.2	1.5
		Advanced	Beginner	5.2	13.2	-8*
		Advanced	Intermediate	5.2	14.7	-9.5*

* The program effect is statistically significant at α = .05.

Appendix D. In-Person SIP and 1-Week SIP

In addition to program evaluation work for SIP and SPP, the study team conducted similar analyses for two other GWC summer programs: an in-person SIP for the 2019 cohort; and a 1-week SIP, a pilot model implemented only in 2020. The following sections summarize the analytic samples and findings for these two programs.

Analytic Sample

For the in-person SIP, the treatment group had a higher percentage of students who were eligible for free or reduced-price lunch, multiracial students, students who were members of historically underrepresented groups, and students who majored in a CS-related field. The comparison group had a higher percentage of students who were not eligible for free or reduced-price lunch and Asian students. For the 1-week SIP, the treatment group had a higher percentage of students who were not eligible for free or reduced-price lunch and students who majored in a CS-related field. The comparison group had a higher percentage of students who were not eligible for free or reduced-price lunch. Exhibit D1 shows the distribution of baseline measures for the treatment and respective comparison students we included for the in-person SIP and 1-week SIP evaluation analyses. We also checked the baseline balances and ensured that treatment and comparison students were equivalent on those baseline measures, with all standardized mean differences between -0.25 and 0.25.

Exhibit D1. In-Person SIP and 1-Week SIP Demographics

	In-person SIP		One-week SIP	
	Treatment (*N* = 1,079)	**Comparison (*N* = 1,830)**	**Treatment (*N* = 866)**	**Comparison (*N* = 1,891)**
Eligible for free or reduced-price lunch				
Yes	55.8%	43.9%	33.0%	38.7%
No or unsure	38.1%	48.3%	58.7%	52.9%
Race/ethnicity				
Asian	29.6%	39.9%	41.2%	43.3%
Black/African American	21.7%	25.5%	18.2%	16.8%
Hispanic/Latinx	21.5%	19.1%	16.9%	17.9%
Multiracial	8.9%	3.8%	3.9%	4.0%
White	14.6%	9.3%	17.6%	15.2%
Additional (small sample sizes)	3.8%	2.4%	2.2%	3.0%
Historically underrepresented students	76.4%	62.8%	53.8%	53.3%
Age – mean (standard deviation)	20.1 (0.8)	20.4 (0.7)	19.1 (0.7)	19.1 (0.7)

	In-person SIP		One-week SIP	
	Treatment (N = 1,079)	Comparison (N = 1,830)	Treatment (N = 866)	Comparison (N = 1,891)
Prior CS knowledge				
No prior knowledge	100%	99.9%	87.9%	85.1%
Beginner	0%	0%	0.6%	0.4%
Intermediate	0%	0.1%	11.5%	13.3%
Advanced	0%	0.1%	0%	1.3%
Grade				
Summer after freshman year	0%	0.3%	0.0%	0.0%
Summer after sophomore year	47.9%	50.4%	50.0%	53.3%
Summer after junior year	52.1%	49.3%	50.0%	46.6%
Summer after senior year	0%	0.1%	0%	0.1%

* The program effect is statistically significant at α = .05.

Program Effects

We examined the percentage of students who entered a CS-related major in the in-person SIP and 1-week SIP, compared to those who were interested in the same programs but did not participate. We found that students who participated in the in-person SIP were significantly more likely to major in a CS-related field, by 17.1 percentage points. Students who participated in the 1-week SIP were also significantly more likely to major in a CS-related field, by 12.8 percentage points (Exhibit D2). We also examined program effects by student group and found that, across most student groups, programs showed consistently positive effects (Exhibits D3 and D4).

Exhibit D2. In-Person SIP and 1-Week SIP Effects

GWC program	Adjusted percentage of majoring in a CS-related field		Program effect	Standard error	p-value	Confidence interval
	Treatment	Comparison				
In-person SIP	45.3	28.2	17.1	1.8	<.001	12.4–21.8
One-week SIP	42.8	30.0	12.8	2.6	.001	6.8–18.8

* The program effect is statistically significant at α = .05.

Exhibit D3. Estimated In-Person SIP Effects by Student Group

Student group	Adjusted percentage of majoring in a CS-related field		Program effect	Standard error	*p*-value	Confidence interval	
	Treatment	Comparison				Low	High
All students	45.3	28.2	17.1	1.8	<.001*	12.4	21.8
Historically underrepresented							
Yes	30.9	13.3	17.6	1.8	.001*	12.6	22.6
No	31.8	15.7	16	3.3	.002*	8.1	23.9
Free/reduced-price lunch							
Eligible	32.2	14.6	17.6	2	.001*	11.9	23.3
Not eligible or unsure	36.4	19.5	16.9	2.2	<.001*	11.6	22.2
Race/ethnicity							
Asian	45.5	29.3	16.2	2.7	.007*	8	24.3
Black or African American	26.3	11.7	14.7	4.6	.019*	3.4	25.9
Hispanic or Latinx	30.2	8.5	21.7	6.6	.033*	3	40.4
Multiracial	31.5	21.1	10.5	6.9	.217	-10.2	31.1
White	36.5	16.5	19.9	3.8	.003*	10.1	29.8
Additional	29.5	9	20.4	11.9	.141	-9.3	50.2

* The program effect is statistically significant at α = .05.

Exhibit D4. Estimated 1-Week SIP Effects by Student Group

Student groups	Adjusted percentage of majoring in a CS-related field		Program effect	Standard error	*p*-value	Confidence interval	
	Treatment	Comparison				Low	High
All students	42.8	30	12.8	2.6	.001*	6.8	18.8
Historically underrepresented							
Yes	32.8	21.6	11.3	2.7	.006*	4.5	18
No	37.2	22.6	14.6	4	0.005*	5.6	23.6
Free/reduced-price lunch							
Eligible	30.1	17.2	12.9	3.3	.014*	4.1	21.8
Not eligible or unsure	39.3	26.6	12.7	3.4	.004*	5.2	20.3
Race/ethnicity							
Asian	47.3	32.6	14.8	3.8	.007*	5.5	24
Black or African American	35.9	20.3	15.6	5.5	.024*	2.7	28.5
Hispanic or Latinx	33	24.4	8.6	6.5	.254	-9	26.2
Multiracial	28.2	31.7	-3.5	7.4	.652	-21.2	14.2

Student groups	Adjusted percentage of majoring in a CS-related field		Program effect	Standard error	*p*-value	Confidence interval	
	Treatment	Comparison				Low	High
White	37.5	20.9	16.6	3.3	<.000*	9.3	23.9
Additional	23.9	33.2	-9.3	14.8	.551	-44.3	25.8
Prior CS knowledge							
No prior knowledge	20.6	9	11.6	3.1	.007*	4.3	18.9
Intermediate	40	19.9	20.1	4.7	.002*	9.6	30.6

* The program effect is statistically significant at α = .05.

Sensitivity Checking

We applied several different approaches to estimate the propensity scores for sensitivity checking purposes and found that the estimated program effects were similar across different methods, indicating the findings are robust (Exhibit D5).

Exhibit D5. Sensitivity Checking for In-Person SIP and 1-Week SIP Effects

GWC program	Sensitivity checking method	Adjusted percentage of majoring in a CS-related field		Program effect	Standard error	*p*-value	Confidence interval
		Treatment	Comparison				
In-person SIP	GLM IPW	45.3	28.6	16.6	1.9	0	11.8–21.5
In-person SIP	NN 1:1 matching	43.6	26.7	16.9	2.2	0	11.5–22.3
One-week SIP	GLM IPW	42.8	30.1	12.6	2.6	.001	6.6–18.6
One-week SIP	NN 1:1 matching	42.4	30	12.4	2.3	.001	7.1–17.8

* The program effect is statistically significant at α = .05.

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**

24250_03/24