Computer Science for English Learners: Final Evaluation Report Narrative

- By: Julian Betts, Dolores Lopez, Dina Polichar, Susan Yonezawa, and Andrew Zau (University of California, San Diego) (Authors listed alphabetically. Yonezawa led the qualitative research and Betts led the quantitative research.)
- February 2026

Executive Summary

CSforEL Year-6 Overview

Implementation Headlines
All three cohorts (2021, 2022, and 2023) were well attended in regards to the summer intensive PD, which represented approximately 40 hours of professional development for each participating teacher. Teachers attended the online professional development from three states. Teachers gained knowledge on three key areas: strategies to analyze language demands of CS lessons/content, strategies to scaffold language while maintaining rigor, and strategies to increase student engagement and collaboration. Key takeaways/innovations: creation of teacher facilitators to co-lead the professional development for 2022 and 2023 and the "lesson planning algorithm" instituted in cohorts 2022 and 2023.

Fidelity-at-a-Glance
For Key Component 1 (summer intensive PD), all three cohorts (2021, 2022 and 2023) met the threshold within the fidelity matrix. For Key Components 2 (year-long PD) and 3 (school leadership network), only Cohort 1 (2021) met the fidelity of implementation thresholds. Cohorts 2 and 3 did not.

Impact Study Design Snapshot
The quantitative evaluation used a randomized controlled trial (RCT) clustered at the school level, designed to examine the impact of the CS for EL professional development program on Ever Secondary English Learner (ESEL) students' enrollment and performance in the AP CSP course compared with schools that did not participate. The evaluation included 51 schools, with 26 randomized into treatment and 25 randomized into the control condition. The primary outcomes studies include AP CSP enrollment, passing the course, College Board exam participation and College Board exam pass rates.

Impact Study Key Findings
For the pooled sample, the students in treated schools exhibited drops of 0.05 in the share enrolling in AP CSP and in enrolling and passing AP CSP relative to the control group. There were no significant effects on taking the College Board Exam or obtaining scores of 3-5 on the College Board exam.

Significant differences emerged across cohorts with the first cohort generally showing no significant impact, the second cohort showing some positive impacts, and the third cohort showing some negative impacts.

Female students showed more negative impacts on AP CSP enrollment and course passage than other students, and similarly students with a home language of Spanish accounted for the negative impacts on enrollment and course passage more than other students.

Forward Look
Based on the observations of the team evaluating the fidelity of implementation, CSforEL provides a carefully developed intervention focused on providing professional learning to high school AP CSP teachers to help them recruit Ever Secondary English Learners into computer science, and then to provide them with language supports. The randomized trial did not show overall positive impacts on recruitment into the course, although this varied across cohorts, and perhaps because of this, no significant results overall resulted for the share of Ever Secondary English Learners taking the relevant College Board exam or getting a score of 3-5 on that exam. One potential mediator for the impact of the program is teacher participation. In the one cohort that exhibited negative impacts on AP CSP enrollment, teacher participation was markedly lower than the other cohorts in the ongoing professional development activities during the first year after the initial summer professional development. Future iterations with new sets of schools may help to resolve the variations in outcomes observed across cohorts of schools in the present study.

Summary Area I — Implementation

Implementation Study – Narrative Frame

The implementation study examined how the CSforEL intervention was carried out across participating districts and schools, and the extent to which administrators, counselors, and teachers engaged with its two central components: the **Equity Audits** and the **teacher professional development (PD)**. This study aimed to understand how consistently the planned activities were enacted and how participants interacted with the intervention.

As outlined in the CSforEL Logic Model (below), the intervention comprised three primary components: (1) the **equity audit**, which took place with administrators, counselors, and teachers to examine systemic barriers to English Learner (EL) participation in computer-science courses; (2) a **week-long summer PD**, designed to build teachers' pedagogical capacity to integrate language scaffolds into AP Computer Science Principles (AP CSP) instruction; and (3) **academic-year CSforEL PD,** which extended teachers' collaboration and reflection throughout the school year. Each component was implemented sequentially across three cohorts, beginning with the equity audits, which started the Spring (2020, 2021, 2023) prior to the summer and academic year PD (2020-2021; 2021-2022; 2023-2024).

The implementation study assessed both the **delivery** and **quality of engagement** within these components. Specifically, analyses explored (a) the degree to which planned activities were completed at the school and teacher levels; (b) the depth of participation among administrators, counselors, and teachers; and (c) how participants applied the knowledge and practices emphasized in the CSforEL model. Quantitative indicators, such as participation records and session attendance, were complemented by qualitative evidence from meeting notes, facilitator observations, teacher reflections, and interview transcripts. Together, these data provided a comprehensive understanding of how the intervention operated in practice, the conditions that supported or constrained its success, and the ways in which schools and educators enacted the program's equity-driven goals.

Equity Audit Implementation: The equity audits (also known as School Leadership Network) were designed to help administrators, counselors, and AP CSP teachers identify systemic barriers affecting English Learners' (ELs') access to and success in computer science courses. Each participating school assembled an equity-audit team of two to three members (typically an administrator, counselor, and CS teacher) who engaged in structured, data-driven reflection and planning activities facilitated by the CSforEL program team.

The audits began in the spring semester preceding the treatment year and continued through the academic year. Teams met virtually for approximately one hour each month to review school data, reflect on findings, and determine next steps. Between meetings, members completed asynchronous activities—such as gathering course-enrollment data, identifying stakeholders for interviews, and reviewing draft resources—and participated in interviews at the beginning and end of the process to discuss evolving perspectives on equity in computer science education.

Analysis of the Equity Audit implementation looked at how closely participating schools carried out the planned activities and the extent of participant engagement throughout the process. Program documentation and facilitator records were reviewed to determine whether core components—such as virtual meetings, asynchronous assignments, and interviews—were completed as intended. Attendance and participation logs were used to assess the overall level and consistency of involvement among school teams. To complement these quantitative indicators, qualitative analyses of meeting notes, participant reflections, and interview transcripts were conducted to assess the depth and quality of engagement. These data illuminated how administrators, counselors, and teachers collaborated to identify institutional barriers and translate equity-focused discussions into actionable plans for expanding English Learner access and success in computer science courses.

In the next section we discuss the Summer and year-long teacher professional development components of the implementation study.

Teacher Professional Development

Summer and Academic Year Professional Development Implementation Study: The teacher PD component of the CSforEL intervention was designed to build AP CSP teachers' capacity to integrate language scaffold supports into computer science instruction. It consisted of two main phases: (1) a **five-day (approximately 40-hour) synchronous summer institute**, and (2) approximately **16 hours of after-school or evening Professional Learning Community (PLC) sessions** conducted throughout the academic year. The PD was facilitated by university and CSTA partners using a **lesson study cycle** (Cerbin & Kopp, 2006; Fernandez, 2002; Rock & Wilson, 2005) and the **SSTELLA framework** (Tolbert et al., 2014), both emphasizing reflective practice, peer collaboration, and the integration of language and computational thinking in CS classrooms.

The Summer Professional Development (PD) served as the foundation of the CSforEL intervention, providing AP CSP teachers from treatment schools with an intensive, collaborative introduction to the program's instructional framework. Conducted over five consecutive days (approximately 20–25 hours), the institute was structured to balance conceptual learning with guided practice and reflection. Each day typically included a progression of activities—beginning with a brief welcome and goal setting, followed by exploration of the day's focus on language demands within computer science, interactive modeling of strategies by facilitators, collaborative planning time with coaches, and structured reflection to consolidate learning. This predictable structure supported continuity across sessions and allowed teachers to incrementally deepen their understanding of both the content and pedagogical dimensions of teaching CS to English Learners. This PD was designed pre-pandemic to be in person, and pivoted during the pandemic to take place online.

The PD providers and participating teachers focused on three overarching components through a combination of real-time discussion, scaffolded activities, provider modeling, and participant practice sessions:

1. Strategies to analyze language demands of CS lessons and content;
2. Strategies to scaffold language while maintaining rigor; and
3. Strategies to increase student engagement and collaboration.

Throughout the summer PD, facilitators guided teachers in analyzing the linguistic complexity of CS content and in developing instructional routines that make computational ideas accessible through purposeful language support. The training emphasized strategies such as interpreting and producing written and spoken language through collaboration, developing academic vocabulary in CS contexts, and designing classroom interactions that promote equitable participation. Teachers engaged in modeling and practicing high-leverage collaborative routines—including pair programming, paraphrase passports, conversation roles, and structured group discussions—each designed to promote meaningful language use, accountability, and shared problem solving among students.

The summer PD's combination of discussion, scaffolded activities, modeling, and participant practice created a learning environment that mirrored the types of collaborative and language-rich classrooms the program sought to foster. By the end of the institute, teachers had developed classroom-ready plans that integrated language scaffolds into their AP CSP instruction and established professional connections that would continue to be reinforced through the academic-year professional learning communities (PLCs)

Analysis of implementation data focused on how well the summer sessions adhered to the intended design and how actively teachers participated and applied new learning. Program records and facilitator notes were reviewed to document the sequence and content of sessions, ensuring that the major instructional components—language analysis, scaffolded practice, and collaborative planning—were implemented as intended. Attendance logs were analyzed to capture overall levels of participation across sites. Qualitative data, including facilitator observations, teacher reflections, and post-institute surveys, provided insight into participants' engagement and perceptions of the training's relevance. These data also helped assess how well teachers understood and were prepared to apply strategies such as purposeful grouping, home-language use, and explicit scaffolding of technical vocabulary in their AP CSP classrooms.

Following the summer PD, teachers engaged in ongoing academic-year PD through PLC meetings held approximately once per month during the academic year. These sessions, facilitated by local CSTA chapters and program staff, provided teachers with structured reflection opportunities aligned to the AP CSP course sequence. Teachers shared implementation experiences, analyzed student work, and discussed adjustments to better support EL participation and language used in computing contexts. Facilitators offered targeted feedback, and teachers collaboratively refined lesson plans to incorporate strategies introduced during the summer PD.

The analysis of the academic-year professional learning communities (PLCs) examined the consistency and depth of implementation across treatment schools and the extent to which teachers continued to apply and refine strategies introduced during the summer institute. Documentation from facilitators, session agendas, and meeting notes was reviewed to assess alignment between the planned sequence of topics and those covered throughout the year. Participation records provided a picture of the regularity of teacher engagement in PLC activities, while qualitative evidence from teacher reflections and interviews illustrated how teachers incorporated CSforEL practices into their instruction.

Overall, findings from the implementation study sought to complement the quantitative results of the impact evaluation by providing contextual understanding of how fidelity, depth of participation, and local adaptation influenced the program's overall effectiveness. Together, these analyses offered a comprehensive picture of how the EIR-funded CSforEL initiative functioned in practice and how its dual focus on systemic equity and teacher learning contributed to advancing English Learners' participation and success in AP Computer Science Principles.

School and Student Demographics

The statistics in this section are derived from Table D.2-1, which also provides many additional demographic characteristics.

The schools in the treatment and control groups on average serve a disadvantaged population. For example, the percentage of students receiving meal assistance in the year before the school's cohort entered the program was 68% for control schools and 69% for treatment schools. For other comparisons we will use shorthand notation such as, for this first comparison, "68% (C) and 69% (T)". The percentage of students who were English Learners was 28% (C) and 27% (T). School enrollment averaged 1588 and 1652 at C and T schools, respectively.

Turning to the ESEL students themselves, their baseline math and ELA test scores, in standardized terms (standardized by grade and year across T and C schools within each Local Education Authority), were about -0.6 for math and -0.8 for ELA. The percentage receiving meal assistance was about 84% in both types of schools. The strong majority of ESEL students were Hispanic (86% C and 90% T), with Asians the second largest racial/ethnic group (4.9% C and 4.7% T). In the year before the program began for the given cohort of schools, the percentage of ESELs who were still English Learners was 83% (C) and 81% (T).

Fidelity of Implementation- Evaluation Overview

1. Logic model for the intervention

The CSforEL Logic Model (see Table 1) describes the key components, mediators, and outcome measures that guided the project evaluation.

Key Components

The intervention included three core components. The first two components focused on the professional development (PD) experiences provided to AP CSP teachers in treatment schools. These included a week-long summer PD followed by ongoing academic-year professional learning**.** Teachers from participating treatment schools were invited to attend the summer PD and to continue engagement through structured professional learning communities during the school year. The PD activities were implemented sequentially across cohorts: Cohort 1 began in Summer 2021, Cohort 2 in Summer 2022, and Cohort 3 in Summer 2023. Each cohort continued participation throughout the corresponding academic year (2021–22 for Cohort 1, 2022–23 for Cohort 2, and 2023–24 for Cohort 3).

The third component centered on the equity audit process offered to each participating high school's administrative, counseling, and CS teaching staff. As described earlier, these equity audits [were also designed to help schools examine systemic barriers to English Learner (EL) participation and success in computer science. The audits began in the spring semester prior to each cohort's summer PD—Spring 2020 for Cohort 1, Spring 2021 for Cohort 2, and Spring 2022 for Cohort 3—and continued throughout the corresponding treatment year (2021–22, 2022–23, and 2023–24, respectively).

Together, these three components represented the essential mechanisms through which the CSforEL project sought to build educator capacity, shift institutional practices, and improve access and outcomes for current and former English Learners in AP Computer Science Principles.

Mediators

Key Components 1 (Summer PD) and 2 (Ongoing PD) were provided to individual participating teachers and were expected to influence two primary mediating factors. First, teachers were anticipated to develop greater capacity to use language scaffolds that made computer science content more accessible to their English Learner (EL) students. Second, they were expected to demonstrate more positive beliefs and attitudes regarding the ability of EL students to successfully learn and engage in computer science.

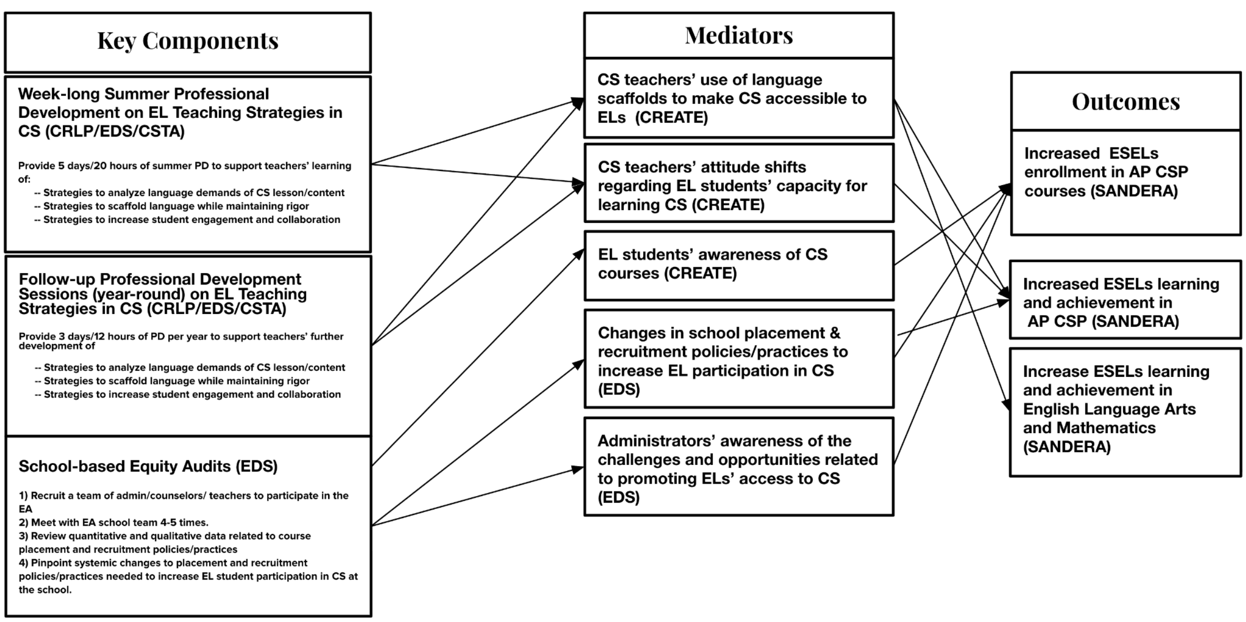
Key Component 3 (Equity Audits) was implemented with school teams composed of teachers, counselors, and administrators who participated in the CSforEL professional development program. This component served as a school-level mechanism intended to influence institutional mediators related to CS recruitment and course placement policies. Through participation in the audits, school staff examined the challenges and opportunities that affected EL and Ever Secondary EL (ESEL) students' enrollment and successful completion of computer science courses. Engagement in these activities was expected to promote changes in administrators' and counselors' approaches to course placement and student advising practices.

As with the CSforEL PD, equity audits were conducted at treatment schools only. Each participating site formed a team that included at least one or more members who attended a series of facilitated audit meetings. The number of sessions attended by school representatives determined whether a site met participation thresholds. During these meetings, program developers collaborated closely with school staff—conducting joint data analyses, facilitating feedback loops, and reviewing school reports—to support data-informed planning and decision-making aimed at expanding equitable access to computer science.

Outcome measures included:

1. Increased ESELs enrollment in AP CSP courses
2. Increased ESELs learning and achievement in AP CSP
3. Increased ESELs learning and achievement in English Language Arts and Mathematics

Table 1. CSforEL Logic Model



Note: In the logic model, acronyms in parentheses refer to the group implementing or evaluating a given activity. On the evaluation side, CREATE (the Center for Research on Education, Assessment and Teaching Excellence, at UCSD) and SanDERA (the San Diego Education Research Alliance at UCSD) led the fidelity of implementation and quantitative evaluation respectively. Implementers and developers of the program are the Computer Science Teachers Association (CSTA) and at UCSD the Statewide California Reading and Literature Program (CRLP) and the Education Studies Department (EDS).

2. Research questions for evaluation of implementation

a. **Implementation**: Were the key components of the **summer professional development** offered to participating teachers implemented as planned? Notably, did participating teachers attend enough (hours and combination of hours) to satisfy threshold requirements of up to 20 hours for implementation to count as meeting the threshold of adequate dosing of the PD intervention?

b. **Implementation**: Were the key components of the ongoing, year-long professional development offered to participating teachers implemented as planned? Notably, did the PD team schedule, hold, and complete ongoing PD with participating teachers of up to 16 hours of PD per teacher?

c. **Implementation**: To what degree did the participating teachers as a group and as individual participants satisfy threshold requirements for BOTH the summer PD and the ongoing academic year PD hours (up to 20 and 16 respectively) to amount to what is considered adequate dosing of the PD to see associated changes in intermediate outcomes?

d. **Scale Up and Replication**: What are formative (annual) and summative (end of grant) understandings about the hindrances and affordances to carrying out successfully the PD implementation at the individual teacher and group levels? Are there specific lessons to be learned from offering professional development virtually as opposed to in-person?

4. Measuring fidelity of implementation

We broke down the primary three key components indicated on the logic model (Table 1) and the fidelity implementation matrix (see Appendix A) to illustrate how they were calculated at the teacher (key components 1 and 2) and school (key component 3) levels to measure thresholds of implementation.

Across the three components, participating schools needed to reach component thresholds at a rate of 50% within the cohort in order for the sample to meet the set threshold to constitute participation in the initiative.

Key Component 1 - Week-long Summer Professional Development on EL Teaching Strategies in CS (CRLP/EDS/CSTA)—calculated at the teacher unit of implementation

Three strategies were introduced to participating teachers during the summer PD experience. These three strategies had been identified by the PD development team for CSforEL as critical to assisting teachers in improving their instructional approaches to English Learners in

Advanced Placement (AP) Computer Science Principles (CSP) courses. These strategies included 1) the ability to analyze language demands of CS lessons and content, 2) the ability of teachers to scaffold language while maintaining rigor in content and instructional approaches, and 3) the development of strategies to increase student engagement and collaboration in the classroom.

In the implementation of the PD for Key Component 1, program designers indicated that teachers had to receive a minimum of 6 of 8 hours of PD in analyzing language demands and 6 of 8 hours of PD in scaffolding language, while maintaining rigor, in order to achieve a reasonable threshold in PD participation. In contrast, PD developers felt that teachers could receive 0 of 4 hours of strategies to increase student engagement and collaboration in the classroom during the summer PD and still meet the set threshold because this topic in particular was revisited and deepened during the ongoing PD provided during Key Component 2.

Key Component 2 - Follow-up Professional Development Sessions (year-round) on EL Teaching Strategies in CS (CRLP/EDS/CSTA)–calculated at the teacher unit of implementation

The same three strategies introduced in Key Component 1 during the week-long summer PD were revisited and deepened in Key Component 2 during the academic year follow-up sessions. However, in the case of Strategy 1, teachers were able to complete 0 of 3 hours of PD because a heavy emphasis had been placed on this particular area of learning in Key Component 1 (summer PD).

For Strategy 2, however, ongoing PD was critical for teachers to gain additional proficiency in this set of strategies (scaffolding language while maintaining rigor). Therefore, participating teachers were required to have attended 4 of 7 hours on this strategy in order to meet threshold requirements within the strategy.

For Strategy 3, increasing student engagement and collaboration, teachers had to attend 4 of the 6 hours of PD on this topic in order to meet threshold requirements within the strategy. This higher threshold for this strategy during the ongoing PD was due to the fact that teachers had been allowed to miss this strategy entirely during the summer PD, where it had been introduced but not covered in depth.

In the implementation of PD for Key Component 2, participating teachers had to attend a minimum of 8 of the 16 hours of PD offered in order to meet thresholds at the individual teacher level in the categories described above.

Key Component 3- School-based Equity Audits (EDS)

Thresholds for the Equity Audit (aka School Leadership Network) participation were measured by the number of meetings attended by a minimum of one Equity Audit school team member meeting with at least one member of the University of California San Diego, Education Studies, Equity Audit program team. There were three primary tasks conducted during Equity Audit meetings: 1) collaborative (planning) meetings, 2) data analysis and feedback meetings, and 3) reception of school-based reports based on gathered Equity Audit team data meetings.

Each of the component parts mentioned above (1 through 3) afforded teams different scores based on their level of participation. At a minimum, teams had to hold two collaborative meetings, participate in one data analysis and feedback meeting, and participate in one meeting where they received their school report from the Equity Audit program designers. If these minimum threshold requirements were met, school teams reached a score level of “3,” which was deemed adequate implementation of the equity audit process.

5. Data collection plan

The fidelity of implementation matrix (see Appendix A) highlighted how indicators for Key Components One, Two, and Three—summer PD, ongoing PD, and Equity Audits—were measured. Fidelity measurements relied on data from the first year of program implementation for each of the cohorts, 2021–2022 (Cohort 1), 2022–2023 (Cohort 2), and 2023-2024 (Cohort 4). Data were collected by the California Reading and Literature Project (CRLP), with assistance as needed from CSTA staff as well as the Equity Audit research team members, at the end of the summer (Key Component One) and academic year (Key Components Two and Three) and were handed off to the evaluation team.

Professional Development Attendance Records: CRLP staff kept track of teachers' daily/hourly attendance for the week-long summer Professional Development (PD) as well as the year-round follow-up PD sessions. Measures for this data source included the number of hours each teacher participated in each of the PD topics addressed: strategies to analyze language demands of CS lesson/content, strategies to scaffold language while maintaining rigor, and strategies to increase student engagement and collaboration (indicators for key components 1 and 2).

Attendance Records of Year-long Equity Audit Meetings: The Equity Audit (EA) research team kept track of the number of meetings it had with members of the school's EA team throughout the school year. Attendance records of collaborative meetings and one-on-one analysis/feedback meetings were used to measure indicators for key component 3 of the fidelity measure.

Online platform: Professional development sessions and equity audit meetings were delivered using an online platform. Log-in data from this platform were used to keep track of teacher attendance and participation in the PDs and EA evaluation meetings. (Teacher friendly platforms like Google forms and Google sheets were used to maximize teacher and school access.)

Agenda and Notes from Summer-long Professional Development (PD): CRLP and CSTA staff kept records of PD agendas and notes that described teacher responses, reactions, challenges, and support strategies discussed. This data source was captured by the two organizations to help identify challenges and guide the direction of forthcoming summer PDs (particularly summer PD in which Cohorts 2 and 3 participated). These data also helped measure teachers' engagement and overall participation in the summer PD across the cohorts.

Professional Development Materials Repository: CRLP and CSTA created an online material repository inclusive of material used during the summer PD (EL strategies, lesson plan, workbooks, resources, etc.) for teachers to use. This material was used to measure topics discussed during the PD sessions and teacher interaction with these items.

Agenda and Notes from Equity Audit (EA) Meetings: Agenda and notes were used to measure which member(s) of the schools' equity audit teams met with the EA evaluation team and how these members participated in the audit process.

Tracker of School Equity Audit Member(s): The Equity Audit evaluation team kept a list of the school's equity audit members and their background (principal, resource teacher, reading specialist, special education, etc.). This information was used as a quantitative indicator that described each school's EA team members.

Professional Development Teacher Tracker: The CRLP (with CSTA) kept a list of the teachers who participated in the summer and ongoing academic year PD. This data included the school's name and the teacher's background (ethnic background, course(s) taught, number of years they've taught, etc.). This information was also used as demographic indicators that described teachers who participated in the PD.

As mentioned initially, the CRLP (with CSTA) and Equity Audit evaluation teams working with Education Studies shared these data sources with the CSforEL evaluation team on a regular basis. These data were used as indicators to help measure the implementation of fidelity for the three key components (summer PD, ongoing/academic year PD, and Equity Audits). The measurement of fidelity was described in the fidelity implementation matrices below (see Appendix A). The fidelity matrix outlined the indicators, data sources, and thresholds that were used to assess whether an indicator was implemented with fidelity on behalf of the teacher and school. A combined score for each key component (across all indicators for that component) was then created for each teacher or school, and an overall fidelity score (a binary indicator) was assigned based on whether the teacher's or school's composite score met the pre-specified threshold for fidelity. This allowed conclusions to be drawn about which components were implemented with fidelity by teachers and schools.

6. Fidelity reporting plan

We described the threshold levels set for each indicator within each of the three key components in order to establish implementation with fidelity of *each.*

For all three key components (and the indicators within them) measurements were taken for the three participating cohorts (2021-22 = cohort 1; 2022-23 = cohort 2; and 2023-24= cohort 3), which corresponded to grant years 3, 4, and 5. Note that grant year 1 (2019-2020) was the project's planning year. Additionally, grant year 2 (2020-21) was a pilot year, which was not included in the implementation or impact models due to COVID-19 adjustments.

Fidelity of implementation threshold scores for the program at the participating teacher schools were reported Spring of 2022 (for cohort 1), Spring of 2023 (for cohort 2), and Spring of 2024 (for cohort 3). Summer PD marked the beginning of the PD period for each year and was "counted" in the year leading into the ongoing academic year PD work. Equity audits began the Spring prior, followed by an implementation period that ran from early Spring 2021 to early Spring 2022 (cohort 1), from early Spring 2022 to early Spring 2023 (cohort 2), and from early Spring 2023 to early Spring 2024 (cohort 3).

At the end of Spring 2024, a three-year combined fidelity of implementation threshold report was compiled for all years of program implementation relevant to the impact evaluation for cohorts 1 and 2. At the end of Spring 2024, a complete fidelity of implementation threshold report was compiled for all program years of implementation for the entire grant period.

Fidelity of Implementation- Key Findings

Below we re-state the four research questions for evaluation of implementation and briefly summarize findings. Appendix B also provides a table with an overview of the full results by key component.

1. Implementation: Were the key components of the summer professional development (Key component 1) offered to participating teachers implemented as planned? Notably, did participating teachers attend enough (hours and combination of hours) to satisfy threshold requirements of up to 20 hours for implementation to count as meeting the threshold of adequate dosing of the PD intervention?

Implementation thresholds for Key Component 1 were met across all three cohorts. To be considered implemented with fidelity, Key Component 1 required that at least 50% of teachers within each cohort meet the individual-level participation threshold. In Cohort 1 (2021–2022), 9 of 14 treatment teachers (64%) met the individual-level threshold. In Cohort 2 (2022–2023), 6 of 7 treatment teachers (86%) met the threshold. In Cohort 3 (2023–2024), 6 of 9 treatment teachers (67%) met the threshold.

2. Implementation: Were the key components of the ongoing, year-long professional development offered to participating teachers implemented as planned? Notably, did the PD team schedule, hold, and complete ongoing PD with participating teachers of up to 16 hours of PD per teacher?

Implementation requirements for Key Component 2 were met only in Cohort 1. This component required participation by at least 50% of teachers within each cohort at the individual level. Only Cohort 1 met the implementation requirements for Key Component 2. Participation at the individual level was achieved by 64% of treatment teachers in Cohort 1 (9 of 14), compared with 43% in Cohort 2 (3 of 7) and 33% in Cohort 3 (3 of 9).

3. Implementation: To what degree did the participating teachers as a group and as individual participants satisfy threshold requirements for BOTH the summer PD and the ongoing academic year PD hours (up to 20 and 16 respectively) to amount to what is considered adequate dosing of the PD to see associated changes in intermediate outcomes?

Overall, participating teachers met the required PD dosage to varying degrees at both the group and individual levels across cohorts. Adequate PD dosage was defined as sufficient participation in both summer PD (up to 20 hours) and ongoing academic-year PD (up to 16 hours) to support expected changes in intermediate outcomes.

At the group level, teachers consistently met participation requirements for summer PD across all cohorts, as evidenced by findings for Key Component 1. In each cohort, more than half of participating teachers completed the minimum hours required at the individual level, indicating adequate exposure to the core PD content delivered during the summer.

In contrast, participation in ongoing academic-year PD was less consistent across cohorts. While Cohort 1 met group-level participation expectations for academic-year PD (Key Component 2), Cohorts 2 and 3 did not reach the minimum level of participation required among teachers as a group. This suggests that, beyond Cohort 1, teachers did not consistently receive the full dosage of sustained, follow-up PD intended to reinforce and extend summer learning.

At the individual level, a subset of teachers in each cohort completed sufficient hours across both summer and academic-year PD to be considered adequately dosed. However, this pattern was strongest in Cohort 1, where a majority of teachers met participation expectations in both PD phases. In later cohorts, fewer teachers accumulated the combined PD hours necessary to ensure consistent reinforcement over time.

In addition, implementation of the equity audit component varied across cohorts. The equity audit was implemented in earlier cohorts and informed PD refinement during those years; however, implementation was limited in Cohort 3, as the third cohort was added beyond the original project scope and implementation timeline. As a result, equity audit findings reflect implementation in earlier cohorts and should be interpreted accordingly.

Taken together, these findings indicate that adequate PD dosage to support intermediate outcomes was achieved most fully in Cohort 1, partially achieved at the individual level in Cohorts 2 and 3, and not consistently achieved at the group level beyond the summer PD phase. These implementation differences should be considered when interpreting variation in intermediate outcomes across cohorts.

4. Scale Up and Replication: What are formative (annual) and summative (end of grant) understandings about the hindrances and affordances to carrying out successfully the PD *implementation at the individual teacher and group levels? Are there specific lessons to be learned from offering professional development virtually as opposed to in-person?*

Formative (Annual) Understandings: Formative evaluation findings highlight several affordances that supported successful PD implementation at both the individual teacher and group levels. Across all three cohorts, CSforEL evaluators provided structured feedback to program implements at three key points during the academic year: following the summer PD session, throughout the lesson study cycles held during the school year, and at the end of the academic year. These recurring feedback loops enabled implements to respond to emerging implementation challenges in a timely manner.

At the individual level, the usage of teacher interviews, PD observations, classroom observations, and student focus groups provided implementers with actionable insights into teachers' experiences with PD content, its pacing, and its applicability inside the classroom. This information supported ongoing refinements to PD activities and it also helped ensure alignment between PD goals and classroom realities.

At the group level, the consistency of feedback across the sites and cohorts supported more coherent PD planning. Following the first year of implementation (2021–2022), evaluators and PD developers engaged in targeted discussions to identify which forms of feedback were most useful for continuous improvement. These conversations informed refinements to observation and interview protocols in subsequent years, allowing evaluators to focus data collection on information most relevant to strengthening PD delivery and responsiveness to teacher and student needs.

A key formative affordance was the ability to gather and synthesize data across multiple state and regional contexts before providing feedback. This helped mitigate the risk of PD design becoming overly tailored to one regional context (e.g., San Diego) and supported more equitable adaptation for teachers in New Mexico and Texas. This cross-site perspective was especially important for understanding group-level implementation patterns and ensuring broader applicability of PD strategies.

Implementation Successes and Affordances: Across all three cohorts, participation in the summer intensive PD was consistently strong. Teachers from three states attended the online summer PD, which provided approximately 40 hours of professional learning per participant. This sustained participation across cohorts indicates that the virtual delivery model effectively supported access and engagement across geographically distributed sites.

The summer PD supported teacher learning in three core areas: analyzing the language demands of computer science lessons and content; scaffolding academic language while maintaining instructional rigor; and increasing student engagement and collaboration. These focal areas were implemented consistently across cohorts, contributing to coherence in PD design over time.

Several programmatic innovations introduced in later cohorts further strengthened implementation. Beginning in the 2022–2023 cohort, experienced teachers served as facilitators and co-led PD sessions, increasing opportunities for peer-to-peer learning and contextualized support. Additionally, the introduction of a structured “lesson planning algorithm” in the 2022–2023 and 2023–2024 cohorts provided teachers with a concrete and replicable tool to support instructional planning and classroom application. These features represent key affordances for scaling and replicating the PD model in future contexts.

<u>Summative (end-of-grant) understandings and Lessons for Replication:</u> Summative findings suggest that successful PD implementation at scale depends on a combination of accessible delivery models, structured opportunities for sustained engagement, and intentional evaluation systems that support continuous improvement. While virtual PD effectively supported broad participation in intensive summer sessions across all cohorts, maintaining consistent engagement over the academic year proved more variable at the group level, underscoring the importance of ongoing support beyond initial access.

Across cohorts, the integration of structured feedback loops, evaluator–implementer collaboration, and refinements to PD design increased the relevance and usability of PD experiences for participating teachers. At the same time, findings highlight the need to balance adaptability with coherence to ensure fidelity while supporting diverse implementation contexts.

For scale up and replication, findings suggest that virtual PD is a viable and effective approach for expanding access, particularly when paired with intentional structures for feedback, facilitation, and instructional support. Hybrid approaches or additional cohort-based supports may further strengthen sustained participation and collective engagement over time. Overall, the CSforEL PD model demonstrates that scalable implementation is best supported by combining flexible delivery with deliberate evaluation-informed refinement, enabling programs to adapt while maintaining core design features across settings.

Summary Area II — Independent Evaluation

Evaluation Design Overview

Two research groups at UC San Diego committed to conducting an independent mixed-methods evaluation. The San Diego Education Research Alliance (SanDERA) at UCSD, headquartered in UC San Diego's Department of Economics, undertook the quantitative evaluation. CREATE at UC San Diego undertook the qualitative evaluation which focused on measuring the fidelity of implementation for the intervention.

Evaluator Independence Statement

The San Diego Education Research Alliance (SanDERA) and the Center for Research of Educational Assessment & Teaching Excellence (CREATE) acted as independent co-evaluators of the *CSforEL* intervention. In this role, they conducted all key aspects of the evaluation. Specifically, SanDERA was responsible for executing the random assignment, collecting and analyzing student outcome data, and estimating and reporting program impacts on the student outcomes while CREATE conducted the qualitative evaluation including observations, interviews, and focus groups. The evaluation findings were not subject to the approval of the project developer/grantee before being reported.

Study Plan Registration (REES)

The study was registered at the Registry of Efficacy and Effectiveness Studies (REES) at https://sreereg.icpsr.umich.edu/sreereg/ as study 7880. There was one significant change to the study after the initial registration: Instead of seeking two cohorts of schools to participate, the project collected three cohorts of schools. In this way the project reached the number of schools planned for in the original power analysis.

Research Questions List

The RCT impact evaluation of CSforEL addressed the following research questions.

Primary (Confirmatory) Research Questions:

Each of the primary outcomes below was measured cumulatively for students enrolled in grades 10 through 12 in the second year of the intervention at each school. For instance, in question 1 below, we measured whether the given student in grades 10 to 12 at the end of the second year after the introduction of the program had enrolled in the AP course. We excluded grade 9 students in year 2 as it is conceivable that they could have been attracted to the school by the

CS for EL program which was initiated in year 1. We also excluded students in grade 12 in year 1 as they were not observed in year 2, and had only one year of exposure to the program.

1. Do high school Ever Secondary English Learner (ESEL) students at schools that participate in the CS for EL program (treatment) have a higher probability of **enrolling in the AP Computer Science Principles course** than high school ESEL students at similar schools that do not participate in the program (control)?
2. Do high school ESEL students at schools that participate in the CS for EL program (treatment) have a higher probability of **passing in the AP Computer Science Principles course** than high school ESEL students at similar schools that do not participate in the program (control)?
3. Do high school ESEL students at schools that participate in the CS for EL program (treatment) have a higher probability of **taking the AP Computer Science Principles exam** than high school ESEL students at similar schools that do not participate in the program (control)?
4. Do high school ESEL students at schools that participate in the CS for EL program (treatment) have a higher probability of **passing the AP Computer Science Principles exam (i.e., scoring 3 or higher)** than high school ESEL students at similar schools that do not participate in the program (control)?

Secondary (Exploratory) Research Questions:

5. Do high school ESEL students at schools that participate in the CS for EL program show **higher gains on statewide ELA tests** than high school ESEL students at similar schools that do not participate in the program (business-as-usual)?

6. Do high school ESEL students at schools that participate in the CS for EL program show **higher gains on statewide Mathematics tests** than high school ESEL students at similar schools that do not participate in the program (business-as-usual)?

7. Does enrollment in **AP Computer Science Principles courses** of students who are non-ESELs change as a result of the program?

8. Is there a difference in any of the program impacts listed above between cohort 1, cohort 2, and cohort 3 schools?

Question 7 emerged during the pilot year (2020-21) when the research team worked with CS teachers to design the Professional Development. Many of the ideas that emerged during this pilot year had to do with providing students at a school with better information to demystify computer science and to dispel myths about computer science, for example by informing students about the wide variety of occupations that make use of computer science skills. This information was shared at treatment schools, and it could have a beneficial side effect of encouraging students who were never ELs to take AP CSP.

Question 8 also emerged during the pilot year when it became clear that the team was adopting a Continuous Improvement strategy that used feedback from teachers to fine-tune program elements such as professional development (PD) for the teachers. If the program team continued to fine-tune the PD from what it learned from working with cohort 1 schools in 2021 through 2023, it could lead to different, perhaps more positive, outcomes for cohort 2 schools that participated in the program in 2022 through 2024, and cohort 3 (participating in 2023 through 2025). Similarly, at the second round of summer PD sessions, newly entering teachers from cohort 2 schools interacted with returning teachers from cohort 1 schools, and the lessons shared by the latter could influence the impact of the program in cohort 2.

For each of these 8 questions, we estimated cumulative program impacts at the end of the second year for students currently in grades 10-12. In these analyses conducted at the end of the second year of the program, we excluded current grade 9 students because theoretically these students could be joiners who endogenously chose to come to the school because of the program one year after its introduction. For the same reason, we excluded in year 2 students in grades 10-12 who enrolled for the first time in the school in that year. These analyses also excluded students in grade 12 in year 1 because unlike the students in grades 9-11 in year 1 they would have only one year of exposure to the program.

It is possible that the effects in year 1 could be different from the effects in year 2. Two factors suggest that the effects could be larger in year 1. First, the summer PD for teachers took place immediately prior to year 1 and might have a larger impact when the PD is fresh in teachers' minds. Second, teachers sometimes move between schools, between districts, or exit teaching altogether. This could lower the impact of the program in year 2.

On the other hand, four factors suggest that the year 2 effects could be larger. First, we measured *cumulative* differences in the outcomes at the end of two years. This gave students an extra year to take AP CSP. Second, it could be that teachers, even though they received the main PD immediately before year 1, could become more effective in applying the PD in year 2. This idea is common in continuous improvement research, where implementers learn from experience and adapt their use of the program elements in a second year. A third reason that the teachers could become more effective in year 2 is that in the summer after year 1 they participated in a workshop where they shared their first-year experiences. This process could have allowed for sharing of best practices and could also have crystallized teachers' understanding of the program elements. Fourth, one important aspect of the CS for EL program is an equity audit in which researchers will work with teams of teachers and administrators at treatment schools to study data on enrollment and success patterns in AP CSP. This work occurred during year 1. Therefore, it could be that any impact of the equity audit on school practices occurred more during year 2 than year 1.

The implementation and research design teams determined that overall it was more important to study cumulative year 2 outcomes. But an interest remained in year 1 outcomes for the reasons listed above. For this reason, we studied the research questions both using cumulative year 2 effects and year 1 effects. But our confirmatory analyses focus on year 2.

Thus we have five primary research questions and three supplementary research questions, and in addition focus on cumulative year 2 outcomes while also studying year 1 outcomes. Table 2 illustrates, showing all the analyses, and distinguishing between primary ("confirmatory") and secondary ("exploratory") analyses. The rows differentiate between primary and secondary research questions, while the columns distinguish between the time period chosen. Our confirmatory analyses consist of the primary research questions measured cumulatively at the end of year 2. Our exploratory analyses, shown by the three quadrants in grey, analyze either the primary research questions at the end of year 1, or the secondary research questions either at the end of year 1 or cumulatively at the end of year 2.

Table 2 Illustration of Confirmatory Analyses and Exploratory Analyses (White Box Indicates Confirmatory Analyses; Gray Boxes Indicate Exploratory Analyses)

	Outcomes Measured Cumulatively at the End of Year 2 of the Intervention	**Outcomes Measured at the End of Year 1 of the Intervention**
Primary Research Questions (1-4)	Confirmatory	Exploratory
Secondary Research Questions (5-8)	Exploratory	Exploratory

Impact Study – Overview

The quantitative evaluation used a randomized controlled trial (RCT) clustered at the school level, designed to examine the impact of the CS for EL professional development program on Ever Secondary English Learner (ESEL) students' enrollment and performance in the AP CSP course compared with schools that did not participate. Targeting high schools in Southern California, Arizona, New Mexico and Texas, the evaluation and implementation teams recruited schools that offered at least one AP CSP course in 2019-20, the project's planning year. In addition for each entering cohort of schools we verified at that time that AP CSP would be offered in the coming year. The same approach was used to recruit schools in two subsequent cohorts. Overall, three cohorts of schools entered into the program, in summer 2021, 2022 and 2023, and the research team gathered and analyzed outcome data over the next two years following the start of treatment, for each cohort. Ultimately the sample included schools from California, New Mexico and Texas.

After allocating schools to blocks of 2, or in one case 3, similar schools, we randomized treatment within each block. AP CSP teachers at roughly half of the schools received professional development (PD) aimed at improving access and opportunity for English learners, while the other half served as control schools and received no such PD. In all, 26 high schools were randomized into treatment, while 25 high schools were randomized into the control group.

The primary outcomes studied included the probability that ESELs enroll in the AP CSP course, pass the course, take the AP CSP exam, and pass the AP exam with a score of 3 or better. Because redesigning AP CSP classes to engage ESEL students could improve English language skill, performance on statewide English Language Arts exams were evaluated as a potential secondary outcome. Similarly, because the field of computer science is related to mathematics and requires some of the same cognitive skills, performance on statewide Mathematics exams was also included as a secondary outcome. These outcomes are reflected in the research questions. A final secondary outcome was the percentage of non-secondary ELs who enrolled in AP CSP. The team hypothesized that there could be positive spillover effects onto these students from the attempts to recruit ESELs into the course.

Primary & Secondary Outcome Measures

As outlined in the Research Questions above, the key primary outcomes are:

1) Enrollment in AP CSP course
2) Passage of AP CSP course
3) Taking the College Board AP CSP Exam
4) Score on the College Board AP CSP Exam, measured as 1 (success) with a score of 3 or higher on the exam, and as 0 if the student obtained a lower score on the exam or did not take the exam.

Secondary outcome variables include scores on statewide assessments in math and English Language Arts, typically in grade 11.

Intervention vs. Matched-Control Condition

Control high schools, like the treatment high schools, offered at least one AP CSP course in the year before the program began in the treatment schools. AP CSP teachers at control schools continued to recruit for and teach the course as they had before, with no additional instruction or support from the CS for EL team.

Study Sample & Cohorts

The final study sample consisted of 51 schools, 26 treated schools and 25 control schools, spread over three cohorts. Cohort 1 (24 schools) participated from summer 2021 through spring 2023. Cohort 2 (9 schools) participated from summer 2022 through spring 2024. Cohort 3 (18 schools) participated from summer 2023 through spring 2025.

Statistical Analysis Plan

For each outcome we estimated a linear model for individual student outcomes while adjusting for matched-pair blocking and clustering at the school level. The key regressor is an indicator for whether the individual student in fall of the first year was enrolled in a treated school.

We included controls for student gender, race, home language spoken if available (both Spanish and other non-Spanish), grade, a measure of socioeconomic status if available across all districts, and school characteristics including percent EL, percent free- and reduced-price lunch, enrollment level, and average grade 8 math and ELA scores of students currently enrolled in the high school. In addition, we controlled for baseline values of each student’s math and ELA test scores from grade 8, and their EL status (current EL versus former EL). If test scores from grade 8 were not available, then we used test scores from grade 7 or if needed grade 6, while adding controls for the grade from which the baseline score was obtained. For missing explanatory variables, we set them to zero while adding a dummy variable to indicate that value was missing.

More statistical details appear in Appendix C.3.

Baseline Equivalence & Attrition Checks

We tested for baseline equivalence between students enrolled in treatment and control schools in year 1 of each cohort's participation. No statistically significant difference emerged for baseline test scores in math and English Language Arts (ELA). An additional 20 tests for baseline equivalence were performed related to student demographics and grade level, and for several schoolwide characteristics. Of these 20, five showed small but significant differences. A second check is to calculate treatment-control differences as a proportion of the population standard deviation. The What Works Clearinghouse (2022) advises that baseline variables do not need to be included as covariates in outcome models if the absolute value of the treatment-control difference, expressed as a share of the standard deviation, is 0.05 or less. It also recommends that any variable for which this difference is between 0.05 and 0.25 of a standard deviation should be included as a control variable. Both test score differences were below 0.05. For the 20 other variables, 16 had differences below 0.05 and 9 had differences between 0.05 and 0.13. We included in the outcome models all of these baseline variables to ensure baseline equivalence criteria were met, and also to increase statistical precision.

A second concern in RCTs is whether attrition is large and in particular unbalanced between treatment and control groups. For the key outcomes, which are cumulative two-year measures of various AP CSP outcomes, all variables conformed to the What Works Clearinghouse (2022) definitions of a low-attrition study.

For the secondary outcomes – capstone test scores in math and English Language Arts – ELA test score attrition met the WWC requirement for low attrition using the WWC's optimistic bound, but the math test score did not meet the low-attrition-rate standard. The results for math achievement must therefore be viewed with caution.

More details for the analyses of baseline equivalence and attrition appear in Appendices D.1 and D.2 respectively.

Student Outcomes — Key Findings

Below we re-state the four primary and four secondary research questions and briefly summarize findings. We also list the Appendix D tables that provide the full results. (Results for the four primary research questions and the first two secondary questions appear in Table D.1-1.)

Primary Research Questions:

*1. Do high school Ever Secondary English Learner (ESEL) students at schools that participate in the CS for EL program have a higher probability of **enrolling in the AP Computer Science Principles course** than high school ESEL students at similar schools that do not participate in the program (business-as-usual)?*

ESEL students in treatment schools had a significantly lower rate of enrollment in AP CSP courses, with a difference in the probability of enrolling (Treatment – Control, or "T-C") of -0.05.

*2. Do high school ESEL students at schools that participate in the CS for EL program have a higher probability of **passing in the AP Computer Science Principles course** than high school ESEL students at similar schools that do not participate in the program (business-as-usual)?*

ESEL students in treatment schools had a significantly lower rate of passing AP CSP courses, with a difference in the probability of enrolling (T-C) of -0.05. Because students were coded as 0 for this outcome if they did not take the course, these first two results imply that students in treatment schools were less likely to take the course, rather than less likely to pass the course conditional on taking it.

*3. Do high school ESEL students at schools that participate in the CS for EL program have a higher probability of **taking the AP Computer Science Principles exam** than high school ESEL students at similar schools that do not participate in the program (business-as-usual)?*

ESEL students in treatment schools exhibited no significant differences from counterparts in control schools in the probability of taking the AP CSP Exam.

*4. Do high school ESEL students at schools that participate in the CS for EL program have a higher probability of **passing the AP Computer Science Principles exam (i.e., scoring 3 or higher)** than high school ESEL students at similar schools that do not participate in the program (business-as-usual)?*

ESEL students in treatment schools exhibited no significant differences from counterparts in control schools in the probability of obtaining a score of 3 or higher on the AP CSP Exam.

Before turning to the secondary research questions it is worth reflecting on the finding that overall students in treated schools were less likely to enroll in, or to pass AP CSP, but they also exhibited no significant differences in the probability of taking the College Board exam or getting a score of 3 or higher on the College Board exam. This difference could arise for several reasons. Perhaps the two more obvious possibilities are:

1) ESELs in treated schools were taught more effectively once in the AP CSP course, accounting for the lack of a treatment-control difference in College Board exam-taking or College Board exam passage;

2) Although treatment schools showed lower rates of AP CSP enrollment, this reduced negative selection of ESELs into those classes.

It is not possible to test either question in a rigorous experiment because the subsample of students taking AP CSP is non-random. But exploratory regression analysis found little support for either hypothesis.[1]

Secondary Research Questions:

*5. Do high school ESEL students at schools that participate in the CS for EL program show **higher gains on statewide ELA tests** than high school ESEL students at similar schools that do not participate in the program (business-as-usual)?*

ESEL students in treatment schools exhibited no significant differences from counterparts in control schools on statewide ELA tests, given either in grade 11 or as end-of-course statewide exams.

*6. Do high school ESEL students at schools that participate in the CS for EL program show **higher gains on statewide Mathematics tests** than high school ESEL students at similar schools that do not participate in the program (business-as-usual)?*

ESEL students in treatment schools exhibited significantly lower math test scores than their counterparts in control schools in schools on statewide ELA tests, given either in grade 11 or as end-of-course statewide exams. The difference was -0.11 standard deviation. However, we reiterate the warning that the lone outcome that did not meet WWC low-attrition standards was standardized math scores, meaning that this result should be viewed with caution.

*7. Does enrollment in **AP Computer Science Principles courses** of students who are non-ESELs change as a result of the program?*

Non-ESEL students exhibited no significant impact of the program on enrollment in AP CSP courses. Table D.1-2 (first column) shows this result, along with other outcomes for non-ESELS, none of which showed significant impacts of treatment.

8. Is there a difference in any of the program impacts listed above between cohort 1, cohort 2, and cohort 3 schools?

Significant differences emerged in program impacts across the three cohorts (Table D.1-3). Consider the four primary research questions related to AP CSP course enrollment, course passage, exam-taking and exam-passing. Cohort 1 showed no significant impact for any of these. Cohort 2 showed a positive impact of treatment for all four outcomes. Cohort 3 showed negative impacts for all four outcomes.

[1] To test the first idea we re-ran the models for AP CSP College Board exam taking, and obtaining a score of 3-5 on that exam, but including only those who took AP CSP. The treatment variable was not significant in either model. To test the idea that treatment schools had less negative selection into AP CSP we re-did the baseline equivalence tests for identical baseline math and ELA test scores between students in treatment and control schools, but including only those who took AP CSP. Baseline equivalence was maintained for ELA test scores. For baseline math scores, the ESELS who took AP CSP showed significantly lower scores in treated than control schools. This suggests the opposite of the above selection hypothesis: students in treated schools became more negatively selected for the subsets enrolled in AP CSP.

Turning to the secondary outcomes related to test scores (Table D.1-3, rightmost columns), Cohort 1 showed a negative treatment effect for math and an insignificant effect for ELA; Cohort 2 showed insignificant effects for either test score; Cohort 3 exhibited negative impacts for both math and ELA test scores.

The bottom of Table D.1-3 also lists control group means for each dependent variable. Cohort 3, the schools of which formed part of a charter school network in the southwestern United States, had markedly higher enrollment in AP CSP courses, with mean enrollment over two years in the course in control schools of 48% of ESEL students, compared to about 3.8% and 2.5% in Cohorts 1 and 2 respectively. It could be that the more widespread taking of AP CSP in cohort 3 control schools is somehow related to the negative treatment impact unique to that cohort.

Subgroup Findings

Our pre-registered plan identified two groupings of students to use to search for variations in program impact.

Table D.1-4 shows models that allowed for different impacts of treatment on ESELs whose home language was Spanish relative to those with another home language. The results show that the negative and significant impacts of treatment on enrollment in AP CSP, and passage of AP CSP, accrue to those ESELs whose home language was Spanish. There was no significant impact for other ESEL students.

Table D.1-5 shows program impacts by gender. None of the impacts are significant for either group. However, the negative impacts on enrollment and course passage shown earlier for the pooled sample do appear to derive more strongly from the female group, and both of these gender differences are statistically significant even though the individual impacts by gender are not statistically significant.

Exploratory Analysis: Outcomes After Year 1

As an exploratory study, Appendix D repeats all of the full-sample tables discussed above, but with outcomes observed after one year. To facilitate readers' ability to compare the cumulative two-year and the one-year results, the latter tables have the same numbering as the former tables, but with a suffix of -YR1.

There are three areas in which results change materially when outcomes are measured after one year rather than two. In Table D.1-1, treatment effects were negative and significant for enrollment in AP CSP when measured cumulatively over two years. But when measured after one year, this treatment effect became insignificant.

The second area in which notable changes arose was for the results by cohort, which can be seen by comparing the results in Table D.1-3 (two year cumulative results) and Table D.1-3-YR1. Recall that the analysis by intervention cohort in Table D.1-3 showed no significant effects for Cohort 1, apart from a negative math impact; Cohort 2 showed a positive and significant impact for all four primary outcomes; Cohort 3, in sharp contrast, showed negative impacts for all four main outcomes and the two secondary test score outcomes. Table D.1-3-YR1 shows the same models but estimated using outcomes after only one year. For Cohort 1 there are meaningful changes with three of the main outcomes – enrollment in, passage of AP CSP, and taking the College Board exam, becoming positive and significant. For Cohorts 2 and 3 results are similar in tenor but in a few cases become insignificant. Further details appear in Appendix D.

The third area where measuring outcome after one year mattered is for the gender subgroups, where the negative impacts for women on enrollment in, and passage of, AP CSP were not significant when measured after two years, but become significant when measured after one year instead of two years. (See Tables D.1-5 and D.1-5-YR1.)

Fuller details on all of the analyses using outcomes after only one year appear in Appendix D.

Summary Area III — Sustainability & Scaling

Lessons Learned & Next-Phase Strategy

Formative evaluation played a central role in shaping implementation across cohorts, and this implementation was undergirded by attention to overall scaling and sustainability over the life of the project. For instance, feedback annually was shared with program staff at several points during the year (after the summer, during the year-long PD, and at regular monthly meetings, when requested). This rhythm created welcome space to address PD and implementation challenges as they arose rather than after the fact. Interviews, observations, and student focus groups gathered during the implementation school year (as well as just after the summer PD) also surfaced practical information about how teachers experienced the PD, including what felt useful, what moved too quickly, and what translated into classroom practice. Looking across states also mattered, year to year, to discern patterns that existed from state to state and those that seemed more state specific.

Implementation results point to steady participation in the virtual summer PD, suggesting that a sustainable model of virtual PD was ascertained over the three-year period. Teachers from the three states completed roughly 40 hours of summer PD/intensive learning, suggesting that the online format lowered barriers to access without sacrificing depth. The PD team also ensured that core instructional priorities remained consistent across cohorts, which helped maintain continuity over time, even as the PD providers shifted from the original designers to teacher facilitators. Indeed, in the later years, the addition of teacher-facilitators and a structured lesson planning tool shifted some of the learning from facilitator-led to peer-supported, giving teachers more concrete ways to apply ideas in their own classrooms. This further suggests that sustainability can be achieved through the curation of additional teacher-leadership across districts and programs.

End-of-grant findings suggest that scale depends on more than reach alone. Virtual delivery made participation possible for a wide range of teachers, but engagement during the school year was less uniform, especially at the group level. Feedback cycles and close collaboration between evaluators and implementers improved the relevance of PD activities, yet the work also highlighted the need to balance flexibility with consistency. For future replication, this project's PD suggests that virtual and/or hybrid approaches, when paired with clear structures for follow-up, facilitation, and ongoing support, can allow programs to adapt to local needs without losing their core design, even across multiple states, districts, types of schools, and contexts.

As noted in the evaluation summary, the program appears well designed. Although the evaluation did not reveal any positive and significant impacts overall, there were variations across cohorts, with the second cohort showing the most promising impacts and the third cohort the least. Future iterations of this program would best be applied to different sets of schools to ensure that impacts are not idiosyncratic to the sample of schools studied here.

Fidelity of implementation was strongest for Cohort 1, the first of three sets of schools to enter the program. We recommend that a focal point for any future extensions of this program should be on finding ways to maintain fidelity of implementation, or put more plainly, ways to ensure that high school teachers continue to participate very actively.

Regarding sustainability, one key issue is whether a program can be extended at lower cost per school than in the early phases of the work. The evaluation team's conclusion is that extension to new schools may be less expensive than the early phases of the current project. We see three reasons for the possibility of falling costs per school: 1) the main work of program design and adaptations has been completed, 2) learning by doing and 3), the selection of some of the most engaged high school teachers to help serve as trainers for subsequent rounds of the program.

Appendices

Appendix A– Fidelity of Implementation Matrix

Indicators	Definition	Unit of implementation	Data Source(s)	Data collection (who, when)	Score for level of implementation at unit level (Teacher)	Threshold for adequate implementation at unit level (Teacher)	Score and threshold for Roll-up to next level -if needed- (cohort level)	Score and threshold for Roll-up to next level -if needed- indicate level	Roll-up to sample level (score and threshold for adequate implementa-tion at sample level)	Expected sample for fidelity measure (n= # units in which the intervention is being implemented)	Expected years of fidelity measurement
Key Component 1 (of 3) - Week-long Summer Professional Development on EL Teaching Strategies in CS (CRLP/EDS/CSTA)											
1. Strategies to analyze language demands of CS lesson/content	Number of hours of Professional Development (PD) provided to a teacher	Teacher	PD attendance records kept by CRLP staff	CRLP staff will provide attendance records to CSforEL evaluation team in the summer, after completion of the week-long PD for each of the cohorts. Summer 2021 = cohort 1 Summer 2022 = cohort 2 Summer 2023 = cohort 3	1= Teacher received 6 of 8 hours of PD developed for language demands 0 = Teachers received less than 6 hours of PD for language demands.	1 = adequate implementation at teacher level				2021-2022 n =14 teachers 2022-2023 n = 7 teachers 2023-2024 n = 9 teachers	2021-22 (cohort 1) 2022-23 (cohort 2) 2023-24 (cohort 3)
2. Strategies to scaffold language while maintaining rigor	# of hours of Professional Development (PD) provided to a teacher	Teacher	PD attendance records kept by CRLP staff	CRLP staff will provide attendance records to CSforEL evaluation team in the summer, after completion of the week-long PD for each of the cohorts. Summer 2021 = cohort 1 Summer 2022 = cohort 2 Summer 2023 = cohort 3	1= Teacher received 6 of 8 hours of PD developed for scaffold language 0 = Teachers received less than 6 hours of PD for scaffold language.	1 = adequate implementation at teacher level				2021-2022 n =14 teachers 2022-2023 n = 7 teachers 2023-2024 n = 9 teachers	2021-22 (cohort 1) 2022-23 (cohort 2) 2023-24 (cohort 3)

3. Strategies to increase student engagement and collaboration	# of hours of Professional Development (PD) provided to a teacher	Teacher	PD attendance records kept by CRLP staff	CRLP staff will provide attendance records to CSforEL evaluation team in the summer, after completion of the week-long PD for each of the cohorts. Summer 2021 = cohort 1 Summer 2022 = cohort 2 Summer 2023 = cohort 3	1= Teacher received 0 of 4 hours of PD developed for engagement and collaboration	1 = adequate implementation at teacher level				2021-2022 n =14 teachers 2022-2023 n = 7 teachers 2023-2024 n = 9 teachers	2021-22 (cohort 1) 2022-23 (cohort 2) 2023-24 (cohort 3)
Key Component Score					Sum of indicators scores at unit level: 1-3	**Teacher level sum score: Adequate Implementation = 3**	NA	NA	**Adequate = 50% or more of teachers in the cohort were at adequate implementation**	2021-2022 n =14 teachers 2022-2023 n = 7 teachers 2023-2024 n = 9 teachers	2021-22 (cohort 1) 2022-23 (cohort 2) 2023-24 (cohort 3)
Fidelity Results					**Threshold**				**50% of teachers in the cohort**	**# of units measured 2021-2022 n =14 teachers 2022-2023 n = 7 teachers 2023-2024 n = 9 teachers**	**Year of Measurement**
					Achieved Score at the Sample Level				**9 of 14 = 64%**	n=14	2021-22
					Met Threshold implemented with Fidelity (Yes, No, N/A)				**Yes**		
					Achieved Score at the Sample Level				**6 of 7 = 86%**	n=7	2022-23
					Met Threshold implemented with Fidelity (Yes, No, N/A)				**Yes**		
					Achieved Score at the Sample Level				**6 of 9 = 67%**	n=9	2023-24
					Met Threshold implemented with Fidelity (Yes, No, N/A)				**Yes**		
Indicators	**Definition**	**Unit of implementation**	**Data Source(s)**	**Data collection (who, when)**	**Score for level of implementation at unit level (Teacher)**	**Threshold for adequate implementation at unit level (Teacher)**	**Score and threshold for Roll-up to next level -if needed- (cohort level)**	**Score and threshold for Roll-up to next level -if needed- indicate level**	**Roll-up to sample level (score and threshold for adequate implementation at sample level)**	**Expected sample for fidelity measure (n= # units in which the intervention is being implemented)**	**Expected years of fidelity measurement**
Key Component 2 (of 3) Follow-up Professional Development Sessions (year-round) on EL Teaching Strategies in CS (CRLP/EDS/CSTA)											

1. Strategies to analyze language demands of CS lesson/content	# of hours of Professional Development (PD) provided to a teacher	Teacher	PD attendance records kept by CRLP	CRLP staff will provide attendance records to CSforEL evaluation team at the end of the academic year 2021-22=cohort 1 2022-23=cohort 2 2023-24=cohort 3	1= Teacher received 0 of 2 hours of PD developed for language demand	1 = adequate implementation at teacher level				2021-2022 n =14 teachers 2022-2023 n = 7 teachers 2023-2024 n = 9 teachers	2021-22 (cohort 1) 2022-23 (cohort 2) 2023-24 (cohort 3)
2. Strategies to scaffold language while maintaining rigor	# of hours of Professional Development (PD) provided to a teacher	Teacher	PD attendance records kept by CRLP	CRLP staff will provide attendance records to CSforEL evaluation team at the end of the academic year 2021-22=cohort 1 2022-23=cohort 2 2023-24=cohort 3	1= Teacher received 4 of 5 hours of PD developed for scaffold language 0 = Teachers received less than 4 hours of PD for scaffold language	1 = adequate implementation at teacher level				2021-2022 n =14 teachers 2022-2023 n = 7 teachers 2023-2024 n = 9 teachers	2021-22 (cohort 1) 2022-23 (cohort 2) 2023-24 (cohort 3)
3. Strategies to increase student engagement and collaboration	# of hours of Professional Development (PD) provided to a teacher	Teacher	PD attendance records kept by CRLP	CRLP staff will provide attendance records to CSforEL evaluation team at the end of the academic year 2021-22=cohort 1 2022-23=cohort 2 2023-24=cohort 3	1= Teacher received 4 of 5 hours of PD developed for engagement and collaboration 0 = Teachers received less than 4 hours of PD developed for engagement and collaboration	1 = adequate implementation at teacher level				2021-2022 n =14 teachers 2022-2023 n = 7 teachers 2023-2024 n = 9 teachers	2021-22 (cohort 1) 2022-23 (cohort 2) 2023-24 (cohort 3)
Key Component Score					Sum of indicators scores at unit level: 1-3	**Teacher level sum score:** **Adequate Implementation = 3**	NA	NA	**Adequate = 50% or more of teachers in the cohort were at adequate implementation**	2021-2022 n =14 teachers 2022-2023 n = 7 teachers 2023-2024 n = 9 teachers	2021-22 (cohort 1) 2022-23 (cohort 2) 2023-24 (cohort 3)
Fidelity Results					**Threshold**				**50% of teachers in the cohort**	**# of units measured 2021-2022 n =14 teachers 2022-2023 n = 7 teachers 2023-2024 n = 9 teachers**	**Year of Measurement**
					Achieved Score at the Sample Level				**9 of 14 = 64%**	n=14	2021-22
					Met Threshold implemented with Fidelity (Yes, No, N/A)				**Yes**		

					Achieved Score at the Sample Level				**3 of 7 = 43%**	n=7	2022-23
					Met Threshold implemented with Fidelity (Yes, No, N/A)				**No**		
					Achieved Score at the Sample Level				**3 of 9 = 33%**	n=9	2023-24
					Met Threshold implemented with Fidelity (Yes, No, N/A)				**No**		
Indicators	**Definition**	**Unit of implementation**	**Data Source(s)**	**Data collection (who, when)**	**Score for level of implementation at unit level (School)**	**Threshold for adequate implementation at unit level (School)**	**Score and threshold for Roll-up to next level -if needed- (All schools at the cohort level)**	**Score and threshold for Roll-up to next level -if needed- indicate level**	**Roll-up to sample level (score and threshold for adequate implementation at sample level)**	**Expected sample for fidelity measure (n= # units in which the intervention is being implemented)**	**Expected years of fidelity measurement**
Key Component 3 (of 3) School-based Equity Audits (EDS)											
Participation by at least one EA team member in collaborative meetings	# of meetings	School	Meeting attendance collected by EA research team members	Evaluator end of academic school year 2021-22=cohort 1 2022-23=cohort 2 2023-24=cohort 3	3 = minimum of 6 collaborative meetings 2 = minimum of 3 collaborative meetings; 1 = minimum of 2 collaborative meetings; 0 = 1 or fewer collaborative meetings	1 = adequate implementation				2021-2022 n =14 schools 2022-2023 n = 7 schools 2023-2024 = 9 schools	2021-22 (cohort 1) 2022-23 (cohort 2) 2023-24 (cohort 3)
Participation by at least one EA team member in one-on-one analysis/feedback meeting with EA researchers	# of meetings	School	Meeting attendance kept by EA research team members	Evaluator end of academic school year 2021-22=cohort 1 2022-23=cohort 2 2023-24=cohort 3	3 = minimum of 3 one-one-one analysis/feedback meetings 2= minimum of 2 one-on-one analysis/feedback meetings; 1 = minimum of 1 analysis/feedback meeting; 0 = 0 one-on-one analysis/feedback meetings	1 = adequate implementation				2021-2022 n =14 schools 2022-2023 n = 7 schools 2023-2024 = 9 schools	2021-22 (cohort 1) 2022-23 (cohort 2) 2023-24 (cohort 3)

EA school teams receive school reports based on EA data.	# of reports received	School	Record of reports received as kept by EA research team members	Evaluator end of academic school year 2021-22=cohort 1 2022-23=cohort 2 2023-24=cohort 3	2 = 2 reports received; 1 = 1 report received; 0 = 0 reports received	1 = adequate implementation				2021-2022 n =14 schools 2022-2023 n = 7 schools 2023-2024 = 9 schools	2021-22 (cohort 1) 2022-23 (cohort 2) 2023-24 (cohort 3)
Key Component Score					Sum of indicator scores at unit level: 0-8	**School-level Sum Score: Adequate Implementation = 3**	NA	NA	**Adequate = 50% or more of schools in the cohort were at adequate implementation**	2021-2022 n =14 schools 2022-2023 n = 7 schools 2023-2024 = 9 schools	2021-22 (cohort 1) 2022-23 (cohort 2) 2023-24 (cohort 3)
Fidelity Results					**Threshold**				**50% of schools in the cohort**	**# of units measured 2021-2022 n =14 schools 2022-2023 n = 7 schools 2023-2024 n = 9 schools**	**Year of Measurement**
					Achieved Score at the Sample Level				**11 of 14 = 79%**	n = 14	2021-22
					Met Threshold implemented with Fidelity (Yes, No, N/A)				**Yes**		
					Achieved Score at the Sample Level				**3 of 7 = 43%**	n = 7	2022-23
					Met Threshold implemented with Fidelity (Yes, No, N/A)				**No**		
					Achieved Score at the Sample Level				**2 of 9 = 22%**	n = 9	2023-24
					Met Threshold implemented with Fidelity (Yes, No, N/A)				**No**		

Appendix B – Fidelity of Implementation Findings By Key Component

Key Component 1 - Week-long Summer Professional Development on EL Teaching Strategies in CS				
Cohort	**Year of Measurement**	**# of Units Measured by Cohort**	**Threshold: 50% of Teachers in the Cohort Achieved Score at the Sample Level**	
1	2021-22	n=14	Achieved Score at the Sample Level	9 of 14 = 64%
			Met Threshold implemented with Fidelity (Yes, No, N/A)	Yes
2	2022-23	n=7	Achieved Score at the Sample Level	6 of 7 = 86%
			Met Threshold implemented with Fidelity (Yes, No, N/A)	Yes
3	2023-24	n=9	Achieved Score at the Sample Level	6 of 9 = 67%
			Met Threshold implemented with Fidelity (Yes, No, N/A)	Yes
Key Component 2 - Follow-up Professional Development Sessions (year-round) on EL Teaching Strategies in CS				
Cohort	**Year of Measurement**	**# of Units Measured by Cohort**	**Threshold: 50% of Teachers in the Cohort Achieved Score at the Sample Level**	
1	2021-22	n=14	Achieved Score at the Sample Level	9 of 14 = 64%
			Met Threshold implemented with Fidelity (Yes, No, N/A)	Yes
2	2022-23	n=7	Achieved Score at the Sample Level	3 of 7 = 43%
			Met Threshold implemented with Fidelity (Yes, No, N/A)	No
3	2023-24	n=9	Achieved Score at the Sample Level	3 of 9 = 33%
			Met Threshold implemented with Fidelity (Yes, No, N/A)	No
Key Component 3 - School-based Equity Audits				
Cohort	**Year of Measurement**	**# of Units Measured by Cohort**	**Threshold: 50% of Schools in the cohort Achieved Score at the Sample Level**	
1	2021-22	n=14	Achieved Score at the Sample Level	11 of 14 = 79%
			Met Threshold implemented with Fidelity (Yes, No, N/A)	Yes
2	2022-23	n=7	Achieved Score at the Sample Level	3 of 7 = 43%
			Met Threshold implemented with Fidelity (Yes, No, N/A)	No
3	2023-24	n=9	Achieved Score at the Sample Level	2 of 9 = 22%
			Met Threshold implemented with Fidelity (Yes, No, N/A)	No

Appendix C – Impact Study Design Supplemental

Below we reproduce the power calculations we made at the time of the study design. Based on this, our goal was a sample of 40 or preferably 50 schools. The final sample is slightly over 50 schools.

We use the all-student standard deviation and mean using data for each district, that is, including both ESELs and non-ESELs, measured in the given outcome year. Minimum detectable effects for planned sample sizes were then produced.

We calculated the MDES using the assumptions listed below in Table C.1.

Table C.1 Main Assumptions Used for a School-Level Cluster Randomized Controlled Trial

Parameter	Value Assumed	Notes on Parameter Choice
Alpha level	0.05	
Two-Tailed or One Tailed Test	Two	
Power	0.80	
Rho (Intercluster Correlation Coefficient)	0.03	
n (Average Number of Ever-EL Students for Level 1)	**300**	Based on the mean number of ELs in the 18 high schools with more than 500 students (2015-16) in one district that committed early to be a participant in the project. We divided this number by 2 because we are focused on those who are still EL's by grade 6, and about half of ELs are reclassified before then. Results are not sensitive to this parameter because number of clusters (schools) predominates.
R_1^2	**0.02**	Proportion of student-level outcomes explained by student characteristics
R_2^2	**0.10**	Proportion of student-level outcomes explained by student characteristics
g*	**5**	Number of school-level covariates used to explain across-school variation (test scores in two subjects, two racial/ethnic measures, % free/reduced-price lunch)

Based on these assumptions, the Minimum Detectable Effect Size (MDES) is as listed in Table C.2 for various choices of the number of schools. (We assume ½ of schools are in the treatment group and ½ are in the control group, so when the number of schools is set to 40 this means 20 schools receiving the program and 20 schools in the control group.)

Table C.2 MDES Versus Number of Schools

Number of Schools	MDES
20	0.235
30	0.186
40	0.159
50	0.141
60	0.128

C.1 Outcome Measures

The four primary outcomes, related to enrollment in AP CSP, passage of AP CSP, taking the associated College Board Exam, and receiving a score of 3-5 on that exam, are all measured cumulatively at the end of the second school year after a cohort entered the program. For the latter three measures, a missing value was coded as 0, to avoid selectivity bias. For example, if a student did not enroll in AP CSP, he or she almost never had a College Board score for the AP CSP exam. The student would then have the binary indicator for passing the College Board exam (with a score of 3 or higher) set to 0.

For the secondary outcomes related to test scores, these test scores were standardized to have mean 0 and standard deviation 1 within grade, year, and Local Education Authority, using all students enrolled in the treatment or control schools in fall of year 1 of participation for the given cohort.

C.2 Comparison School Selection Procedure

Local Education Authorities: Schools in Local Education Authorities (LEAs) in three states in the southwestern United States were selected based on a combination of factors, including an existing program of AP CSP and a large population of English learners. Two LEAs were included based on previously established research-practitioner partnerships as well as an existing program of AP CSP and a large population of English learners.

Schools: Within each participating LEA, schools that offered AP CSP were grouped into pairs matched on school demographic variables during the baseline year, including percent English learners, rates of EL enrollment in AP CSP, standardized test scores, and school size. In one case a trio of similar schools in a given district were placed in a randomization group as well. Within each pair, we randomly selected one of the schools to be recruited to participate in the study and the other were designated a control school. For the single group of three schools, two schools were randomly assigned to treatment. (Because separate intercepts are included for each randomization group, observations were not weighted.)

Teachers: Recruitment efforts targeted the AP CSP teacher at each study school. It was anticipated that there would only be one AP CSP teacher at any given school. However, if any study school had more than one teacher of AP CSP, all teachers were included in the study.

Classes: All AP CSP classes at each study school were included in the study. However, in the majority of cases only one class was offered at each school.

Students: Students were included in the study if they were designated an English Learner at baseline, or if their designation to fluent English proficient occurred in grade 6 or above; we refer to these students as Ever Secondary ELs (ESELs). Furthermore, students were included if they were enrolled in grades 9-11 in fall of the first year of participation in the program for a given pair or trio of schools. Students were included if they later moved to other schools. Students arriving after fall of year 1 were not included, to avoid migration to treatment or control schools that could have been influenced by the existence of the CSforEL program.

C.3 Impact Model Specification

We approach the analysis as a school-level RCT with clustering at the school level, and analysis at the level of the individual student. Within each district, a matching process produces one or more blocks consisting of pairs of high schools, one of which is randomized into treatment. (In one case a trio of schools in a district were matched, with two of three randomly selected for treatment.) Let Y_{istb} denote an outcome for student i currently enrolled in high school s in year t and in block (school pair) b. We denote year t as the first year of treatment and t+1 as the second year of treatment. The key explanatory variable is $TREAT_s$, a binary variable equaling 1 if school s is offered the CS for EL treatment, and 0 otherwise. Let X_i denote a vector of student characteristics that we include as explanatory variables to increase precision. These are measured as of grade 8 to eliminate the possibility that student characteristics could be endogenous with respect to treatment. (We provide further details later.) Let $SCHL_{s,t-1}$ denote a vector of school characteristics measured based on enrollment in year t, but measured as a mean calculated based on individual students' characteristics measured in year t-1, the base year.

After pooling data across all blocks of schools across Local Education Authorities, we estimate the following model:

$$(1)\ Y_{istb} = \sum_{j=1}^{B} \alpha_j + \beta TREAT_s + X_i \gamma + SCHL_{s,t-1} \Lambda + \epsilon_{istb}$$

On the right-hand side of the model, the terms, in order, are fixed effects (dummy variables) for each pair of schools in a block, where there are B blocks in total, and the next three terms are the treatment indicator, a vector of student controls, and a vector of school controls, each multiplied by corresponding coefficients to be estimated, including the coefficient β, and the vectors of coefficients Γ and Λ. The final term is an error term. We report clustered standard errors, with clustering at the level of the school attended in fall of the first year of participation of the given cohort of schools in CSforEL. Clustering allows both for heteroskedasticity across schools and any correlations between any two observations within each school. The interpretation of the key coefficient β is that being offered the CS for EL program causes the outcome variable to change by β units. Thus, it is an intent-to-treat estimator.

Our outcomes are binary indicators that are set to 1 if the student has met a given criterion, and zero otherwise. We include all Ever Secondary EL (ESEL) students in each calculation. For example, research question 3 asks whether an ESEL student has passed the AP exam for AP CSP with a score of 3 or higher. To avoid selectivity bias, this indicator is set to 0 not only for those who take the exam and receive a score below 3, but also for those who take AP CSP but do not take the exam, and those who do not take AP CSP.

Most of our outcomes are specified in the form "Has the student ever taken AP CSP?" or "Has the student ever taken the College Board AP CSP exam, and scored 3 or higher?". The advantage of this cumulative approach of asking if the student has "ever" met a condition is that it overcomes problems of selectivity bias that would result if one instead looked at annual outcomes such as "Did the student take AP CSP this school year?" which would have to drop those students who had already taken AP CSP in a prior year. Nonetheless, our choice to measure outcomes in this cumulative way raises the issue that some students will have taken AP CSP before treatment begins. The use of control group schools controls for this issue. Further, we include as a covariate the percentage of ESELs who had completed AP CSP by year t-1, where year t is the first year of the treatment. This will control for any random differences between treatment and control schools in the prior rate of AP CSP completion.

Subgroups

We performed subgroup analyses for the ESEL group, based on gender, and on whether Spanish is the home language. One of the secondary research questions listed earlier (#8) is: "Is there a difference in any of the program impacts listed above between cohort 1, 2 and 3 schools?" This too is really a subgroup analysis.

We perform these analyses in models that use the entire sample, allowing for tests of different effects between subgroups. We pre-committed to perform power analyses ex ante and to avoid estimating effects for subgroups that are so small that the projected MDES is above 0.4.

For example, to test whether the causal impact of the program differs by gender, we re-estimate (1) as follows:

(2)

$$Y_{istb} = \sum_{j=1}^{B} \alpha_j + \beta TREAT_s(1 - Female_i) + \delta TREAT_s(Female_i) + X_i\gamma + SCHL_{s,t-1}\Lambda + \epsilon_{istb}$$

Here, we allow the impact of treatment to vary between female and all other students. We report a p-value for a test that the two impacts are identical (□=δ). Similarly, we run variants of (1) that allow for the treatment effect to vary by race/ethnicity. In such cases where there are multiple groups, we report on an F-test for equality of the treatment effect across groups.

C.4 Baseline Equivalence Model Specification

The baseline equivalence tests were performed by estimating simplified versions of (1) above in Appendix C.3, in which the given baseline student characteristic was regressed on an indicator for school treatment, dummies for blocks, and with clustering of the error terms at the school level.

Appendix D – Impact Study Findings Tables

D.1 Impact Findings Tables

Table D.1-1 The Impact of Treatment on ESEL Students in CSforEL Schools

	Main Questions:				**Secondary Questions:**	
	Enrolled AP CSP	**Passed AP CSP**	**Took AP CSP Exam**	**Scored 3-5 on AP Exam**	**Grade 11 Math Z-Score**	**Grade 11 ELA Z-Score**
Treatment	-0.0515*	-0.0546*	-0.0239	-0.0033	-0.1076**	-0.0187
	(0.022)	(0.022)	(0.021)	(0.003)	(0.035)	(0.026)
N	12319	12319	12319	12319	4644	5486
R-squared	.351	.349	.428	.104	.324	.455
Dep. V. Mean, Control Group	.120	.114	.107	.0105	-.533	-.606

Robust standard errors in parentheses.
* p<0.05, ** p<0.01

The above table shows the impacts for the four primary outcomes, plus the two secondary outcomes related to test scores in the two rightmost columns. The row labeled "Dep. V. Mean, Control Group" shows the mean of the dependent variable for the control group.

The pre-registered evaluation plan included a test for whether the program increased AP CSP enrollment for non-ESEL students. That result for this secondary research question is shown in the first column below. For completeness, we show the analysis of all outcomes for this group. No impacts are statistically significant.

Table D.1-2 The Impact of Treatment on Non-ESEL Students in CSforEL Schools

	Enrolled AP CSP	**Passed AP CSP**	**Took AP CSP Exam**	**Scored 3-5 on AP Exam**	**Grade 11 Math Z-Score**	**Grade 11 ELA Z-Score**
Treatment	-0.0324	-0.0319	-0.0204	0.0026	-0.0015	-0.0325
	(0.018)	(0.017)	(0.015)	(0.005)	(0.054)	(0.032)
N	37373	37373	37373	37373	17363	18094
R-squared	.162	.176	.225	.051	.465	.399
Dep. V. Mean, Control Group	.0752	.0686	.0693	.0248	.181	.222

Robust standard errors in parentheses.
* p<0.05, ** p<0.01

Table D.1-3 The Impact by Intervention Cohort of Treatment on ESEL Students in CSforEL Schools

	Enrolled AP CSP	Passed AP CSP	Took AP CSP Exam	Scored 3-5 on AP Exam	Grade 11 Math Z-Score	Grade 11 ELA Z-Score
Treatment * Cohort 1	0.0103	0.0120	0.0165	-0.0017	-0.0866*	0.0686
	(0.020)	(0.019)	(0.021)	(0.004)	(0.043)	(0.051)
Treatment * Cohort 2	0.0602**	0.0571**	0.0649***	0.0064**	0.0669	-0.0436
	(0.018)	(0.018)	(0.017)	(0.002)	(0.035)	(0.048)
Treatment * Cohort 3	-0.2081***	-0.2163***	-0.1396**	-0.0130**	-0.5530***	-0.0925*
	(0.051)	(0.048)	(0.040)	(0.005)	(0.158)	(0.036)
N	12319	12319	12319	12319	4644	5486
R-squared	.362	.362	.434	.105	.327	.456
p-value Equal Effects	1.15e-06	7.07e-07	2.68e-06	.00127	.000163	.0721
Control Mean, Cohort 1	.0384	.0323	.0162	.00289	-.622	-.803
Control Mean, Cohort 2	.0254	.0215	.0182	.000651	-.464	-.467
Control Mean, Cohort 3	.482	.472	.489	.0456	.435	-.115

Robust standard errors in parentheses.
* p<0.05, ** p<0.01, *** p<0.001

Table D.1-4 The Impact of Treatment on ESELS by Home Language Spoken

	Enrolled AP CSP	**Passed AP CSP**	**Took AP CSP Exam**	**Scored 3-5 on AP Exam**	**Grade 11 Math Z-Score**	**Grade 11 ELA Z-Score**
Treatment * Spanish Language at Home	-0.0673**	-0.0703**	-0.0350	-0.0040	-0.1086*	-0.0207
	(0.025)	(0.024)	(0.023)	(0.003)	(0.044)	(0.028)
Treatment * Non-Spanish Language at Home	-0.0107	-0.0143	0.0047	-0.0013	-0.1047*	-0.0112
	(0.019)	(0.019)	(0.019)	(0.003)	(0.046)	(0.049)
N	12319	12319	12319	12319	4644	5486
R-squared	.352	.351	.428	.104	.324	.455
Dep. V. Mean, Control Group	.12	.114	.107	.0105	-.533	-.606
p-value Equal Effects	.00239	.00289	.0156	.322	.952	.851

Robust standard errors in parentheses.
* p<0.05, ** p<0.01

Table D.1-5 The Impact of Treatment on ESELS by Student Gender

	Enrolled AP CSP	Passed AP CSP	Took AP CSP Exam	Scored 3-5 on AP Exam	Grade 11 Math Z-Score	Grade 11 ELA Z-Score
Treatment * Female	-0.0332	-0.0365	-0.0171	-0.0059	-0.0066	-0.0071
	(0.021)	(0.020)	(0.021)	(0.003)	(0.037)	(0.038)
Treatment * Non-Female	0.0082	0.0063	0.0091	0.0007	0.0530	0.0219
	(0.012)	(0.011)	(0.014)	(0.004)	(0.105)	(0.092)
N	12319	12319	12319	12319	4644	5486
R-squared	.348	.346	.427	.105	.323	.455
Dep. V. Mean, Control Group	.12	.114	.107	.0105	-.533	-.606
p-value Equal Effects	.0146	.00951	.105	.152	.549	.78

Robust standard errors in parentheses.
* p<0.05, ** p<0.01

Note: A very small number of students were non-binary and were included in the non-female group.

Robustness Checks Using Year 1 Outcomes

This section repeats the above tables but instead measures each outcome after one year of the program.

Table D.1-1-YR1 Robustness Check Using Outcomes Measured After Year 1: The Impact of Treatment on ESEL Students in CSforEL Schools

	Main Questions:				**Secondary Questions:**	
	Enrolled AP CSP	**Passed AP CSP**	**Took AP CSP Exam**	**Scored 3-5 on AP Exam**	**Grade 11 Math Z-Score**	**Grade 11 ELA Z-Score**
Treatment	-0.0358	-0.0370*	0.0065	0.0019	-0.1393**	-0.0259
	(0.018)	(0.018)	(0.006)	(0.001)	(0.040)	(0.025)
N	12319	12319	12319	12319	2431	3251
R-squared	.282	.277	.108	.0317	.369	.528
Dep. V. Mean, Control Group	.0678	.0642	.0128	.00162	-.506	-.55

Robust standard errors in parentheses.
* p<0.05, ** p<0.01

Compared to Table D.1.1, the two of the four confirmatory outcomes at the left of the table that were negative and significant were enrollment in, and passage of, AP CSP. Both of these effects remain negative in Table D.1-1-YR1 when measuring outcomes after one year, but only the passage effect remains significant. Among the two secondary questions related to test scores in Table D.1.1, the math impact remains negative and significant.

Table D.1-2-YR1 Robustness Check Using Outcomes Measured After Year 1: The Impact of Treatment on Non-ESEL Students in CSforEL Schools

	Enrolled AP CSP	Passed AP CSP	Took AP CSP Exam	Scored 3-5 on AP Exam	Grade 11 Math Z-Score	Grade 11 ELA Z-Score
Treatment	-0.017	-0.0164	0.0078	0.0025	0.0225	-0.0639
	-0.013	-0.012	-0.005	-0.002	-0.062	-0.033
N	37374	37374	37374	37374	8532	9250
R-squared	0.147	0.156	0.0444	0.0243	0.496	0.43
Dep. V. Mean, Control Group	0.0387	0.0353	0.0196	0.00948	0.18	0.229

Robust standard errors in parentheses.
* p<0.05, ** p<0.01

The secondary outcomes for non-ESEL students in Table D.1-2 remain insignificant when re-estimated using outcomes measured after one year, as shown in Table D.1-2-YR1.

Table D.1-3-YR1 Robustness Check Using Outcomes Measured After Year 1: The Impact by Intervention Cohort of Treatment on ESEL Students in CSforEL Schools

	Enrolled AP CSP	Passed AP CSP	Took AP CSP Exam	Scored 3-5 on AP Exam	Grade 11 Math Z-Score	Grade 11 ELA Z-Score
Treatment * Cohort 1	0.0357*	0.0351*	-0.0221*	-0.0013	-0.0408	0.0546
	(0.014)	(0.014)	(0.010)	(0.001)	(0.057)	(0.054)
Treatment * Cohort 2	0.0394*	0.0357	0.0208**	0.0043**	0.0254	-0.0362
	(0.019)	(0.019)	(0.007)	(0.001)	(0.049)	(0.067)
Treatment * Cohort 3	-0.1716***	-0.1715***	0.0237	0.0031	-0.5798**	-0.0848*
	(0.036)	(0.036)	(0.015)	(0.002)	(0.191)	(0.037)
N	12319	12319	12319	12319	2431	3251
R-squared	.296	.292	.111	.032	.373	.529
p-value Equal Effects	.0000316	.0000308	.00091	.00842	.0285	.194

Robust standard errors in parentheses.
* p<0.05, ** p<0.01, *** p<0.001

Recall that the analysis by intervention cohort in Table D.1-3 showed no significant effects for Cohort 1, apart from a negative math impact; Cohort 2 shows a positive and significant impact for all four primary outcomes; Cohort 3, in sharp contrast, showed negative impacts for all four main outcomes and the two secondary test score outcomes. Table D.1-3-YR1 shows the same models but estimated using outcomes after only one year. For Cohort 1 there are meaningful changes while for Cohorts 2 and 3 results are similar in tenor but in a few cases become insignificant.

In more detail, for Cohort 1 the impacts of treatment on enrollment in and passage of AP CSP become positive and significant, while the impact on taking the AP Exam become negative and significant. None of these effects is significant when measured cumulatively over two years. For Cohort 2, where all four main outcomes measured at the end of two years were positive and significant, when instead measuring outcomes after one year, this pattern remains, except that the effect on course passage, while still positive, is no longer significant. For Cohort 3, where all impacts in Table D.1-3 were negative and significant, in two cases, taking the College Board exam, and scoring a 3 or higher, the effects become statistically insignificant.

Table D.1-4-YR1 Robustness Check Using Outcomes Measured After Year 1: The Impact of Treatment on ESELS by Home Language Spoken

	Enrolled AP CSP	Passed AP CSP	Took AP CSP Exam	Scored 3-5 on AP Exam	Grade 11 Math Z-Score	Grade 11 ELA Z-Score
Treatment * Spanish Language at Home	-0.0467*	-0.0477*	0.0086	0.0019	-0.1623**	-0.0388
	(0.020)	(0.020)	(0.007)	(0.001)	(0.049)	(0.027)
Treatment * Non-Spanish Language at Home	-0.0075	-0.0095	0.0013	0.0019	-0.0718	0.0248
	(0.013)	(0.013)	(0.008)	(0.001)	(0.055)	(0.054)
N	12319	12319	12319	12319	2431	3251
R-squared	.284	.279	.108	.0317	.369	.528
Dep. V. Mean, Control Group	.0678	.0642	.0128	.00162	-.506	-.55
p-value Equal Effects	.0066	.0079	.316	.977	.198	.218

Robust standard errors in parentheses.

* p<0.05, ** p<0.01

Results when analyzed separately for ESELs whose home language was Spanish, versus another language, were similar in sign and significance when repeated after measuring outcomes after one year: Spanish speakers account for the overall negative impacts on enrollment in, and passage of, AP CSP.

Table D.1-5-YR1 Robustness Check Using Outcomes Measured After Year 1: The Impact of Treatment on ESELS by Student Gender

	Enrolled AP CSP	Passed AP CSP	Took AP CSP Exam	Scored 3-5 on AP Exam	Grade 11 Math Z-Score	Grade 11 ELA Z-Score
Treatment * Female	-0.0297*	-0.0314*	0.0041	0.0017	-0.0299	-0.0198
	(0.015)	(0.014)	(0.006)	(0.002)	(0.064)	(0.040)
Treatment * Non-Female	-0.0010	0.0011	-0.0060	0.0008	0.1168	0.0979
	(0.010)	(0.009)	(0.013)	(0.002)	(0.107)	(0.141)
N	12319	12319	12319	12319	2431	3251
R-squared	.281	.276	.108	.0317	.367	.528
Dep. V. Mean, Control Group	.0678	.0642	.0128	.00162	-.506	-.55
p-value Equal Effects	.0379	.0142	.48	.741	.182	.418

Robust standard errors in parentheses.
* p<0.05, ** p<0.01

Table D.1-5 revealed some differences by gender in treatment effects for enrollment in, and passage of, AP CSP, with lower impacts for females than non-females. However, no impacts for either group were significant. When this analysis was repeated using year one outcomes, that pattern remained. However, the negative impacts for females on enrollment in, and passage of, AP CSP now become statistically significant when these outcomes are measured after one year.

D.2 Baseline Equivalence Tables

Table D.2-1 shows results of regressions of baseline student characteristics on an indicator for treatment in the school year before each cohort of school began the program. (Appendix section C.4 provides further details.) We include two measures of baseline achievement and multiple demographic measures.

Table D.2-1 shows the coefficient for treatment, which measures the difference between the treatment and control groups, using asterisks for significant differences. Other columns show the standard error, the control group mean, the combined sample standard deviation, sample size and R^2. Neither baseline test score differs significantly between the treatment and control groups. Among the baseline demographic variables and, at the bottom, the mean baseline values calculated for the entire school, most differences are not significant. The exceptions are grade in which the student is enrolled, home language being Spanish or another language, an indicator for students who had already taken AP CSP by the baseline year, and the grade of reclassification for former English Learners. The What Works Clearinghouse (2022, page 53) guidelines for checking baseline equivalence are that differences should ideally be less than 0.05 of a standard deviation for a study not to condition on baseline covariates,. Alternatively, WWC specifies that differences between 0.05 and 0.25 of a standard deviation can meet baseline equivalence by controlling for these variables. Most differences are below 0.05 but 9 of the 22 variables have differences between 0.5 and 0.13 of a standard deviation. The largest deviation is for the Hispanic indicator, which is 0.04 higher in the treatment group, or 0.13 of a standard deviation. No variable exceeds the 0.25 limit. Out of an abundance of caution, we include controls for all of these baseline variables in all outcome models.

Table D.2-1 Regression-Based Test for No Baseline Difference Between Treatment and Control Schools

Variable	Treatment Difference	Standard Error	Control Group Mean	Combined Sample S.D.	N	R Squared	Difference as Proportion of a S.D.
Baseline Math Z-Score	-0.0226	(0.025)	-0.596	0.866	4999	.136	-0.026
Baseline ELA Z-Score	-0.0183	(0.036)	-0.762	0.883	4918	.154	-0.021
Grade	-0.0356*	(0.015)	9.129	0.802	8603	.0533	-0.021
Female	0.0135	(0.009)	0.442	0.498	10926	.00451	0.027
Home Language is Spanish	0.0270*	(0.013)	0.713	0.453	10926	.576	0.034
Home Language is Other Than Spanish	-0.0208*	(0.008)	0.227	0.400	10926	.737	-0.052
Meal Assistance	0.0076	(0.006)	0.837	0.365	9119	.188	0.021
White	-0.0040	(0.003)	0.040	0.186	10921	.0195	-0.022
Black	-0.0224	(0.013)	0.033	0.148	10921	.0952	-0.151
Hispanic	0.0423	(0.024)	0.859	0.327	10921	.0895	0.129
Asian	-0.0055	(0.011)	0.049	0.208	10921	.0479	-0.026
Native American	-0.0087	(0.007)	0.013	0.107	10921	.0782	-0.081
Pacific Islander	0.0002	(0.000)	0.001	0.029	10921	.00328	0.007
Other Race or Ethnicity	-0.0019	(0.001)	0.005	0.073	10921	.0197	-0.026
English Learner	-0.0214	(0.012)	0.827	0.404	10926	.123	-0.053
Had Already Taken AP CSP	-0.0044*	(0.002)	0.007	0.071	12319	.0145	-0.062
Reclassification Grade	-0.1289*	(0.050)	9.539	1.289	3342	.172	-0.100
School Enrollment	63.5768	(107.28)	1588.45	684.40	12319	.83	0.093
School Mean Grade 8 Math Z-Score	-0.0054	(0.016)	-0.048	0.110	11442	.805	-0.049
School Mean Grade 8 ELA Z-Score	-0.0052	(0.009)	-0.058	0.107	11442	.912	-0.049
School Share Meal Assistance	0.0090	(0.009)	0.683	0.192	10395	.968	0.047
School Share English Learner	-0.0124	(0.008)	0.277	0.199	12319	.971	-0.062

Robust standard errors in parentheses.
* p<0.05, ** p<0.01, *** p<0.001

D.3 Attrition Findings

i) Cluster-Level Attrition

Schools were grouped into clusters in spring of the school year before the school year of program implementation for a given cohort. Of the 28 clusters created, 3 of 28 or 11% were dropped. In all three cases the drops came about because of a failure to complete a Memorandum of Agreement with the local authority. These drops occurred before the start of the school year of program implementation.

ii) Risk of Bias Due to Joiners

Joiner risk refers to the possibility that news of the program at treatment schools could cause students to transfer to (or from) these schools in order to take part in the programs. To limit joiner bias, the sample consisted only of students in grades 9 through 11 in fall of year 1 of implementation. Students arriving later in that school year or in year 2 were not included for either treatment or control schools. The only possible way that joiner bias could have occurred is if families had discovered the new program in the summer before the start of classes and either moved or used any existing school choice provisions to transfer their student to the treatment school. Such transfers are unlikely given our understanding that news about the program offered to AP CSP high school teachers in treatment schools was not widely circulated during that period..

iii) Individual-Level Attrition

The third possibility is that attrition takes place at the level of the individual students in treatment and control schools.

The main outcomes are cumulative outcomes measured at the end of two school years after the reform is introduced to each set of schools in a treatment/control group. We present two estimates of attrition, the first treating any student with at least one year of outcome data as not missing the outcome, and a second more conservative version that treats a student with less than two years of data observed as missing the cumulative two-year outcome.

Table D.3-1 adopts the first approach. It shows the percentage of treatment and control observations for which each outcome is observed, with the original number of observations in year 1 of the intervention in the first row. In interpreting this table, recall that for outcomes such as obtaining a score of 3-5 on the AP CSP College Board Exam, students who never took the Exam are coded as zero, as are those who took the exam but did not receive a high score. Similarly, the indicator for receiving a passing AP Grade is not set to missing if the student did not take the course, but rather to 0.

We have outcomes for most variables for 100% of the treatment and control groups. The two exceptions are the grade 11/End-of-Course Math and ELA scores, for which on average only 46.7% and 55.2% of students have observations.

Table D.2-1 Percentage Availability of Cumulative Outcomes by Treatment and Control Group

Sample	Treatment	Control	Total	Differential Attrition (T-C)
Full Year 1 Sample (#):	6137	6182	12319	
Percentages Available:				
Enrolled AP CSP	100.0	100.0	100.0	0.0
Passed AP CSP	100.0	100.0	100.0	0.0
Took AP CSP Exam	100.0	100.0	100.0	0.0
Scored 3-5 on AP Exam	100.0	100.0	100.0	0.0
AP CSP Test Score	100.0	100.0	100.0	0.0
AP CSP Grade	100.0	100.0	100.0	0.0
Full Year 1 Sample (#) for Test Scores:	4791	5146	9937	
Percentages Available:				
Grade 11/End-of-Course Math Z-Score	49.0	44.7	46.7	4.3
Grade 11/End of Course ELA Z-Score	56.4	54.1	55.2	2.3

Note: In the bottom panel, the results for Math and ELA test scores are for restricted samples that exclude students in grade 9 in year 1 of the program in states with a Grade 11 test. They are excluded because these students would not have reached grade 11 during the two years in which the program operated.

The attrition rates for the two types of test scores is less than implied in the above table. Two factors contribute to the low availability rate for test scores. In states with mandated testing in grade 11, students who were in grade 9 at the start of the intervention would not have reached grade 11 by the end of two school years. The bottom panel already adjusts for this by removing these students from the calculations. Second, in one state there was no grade 11 exam but rather an End-of-Course exam for students completing a designated course in math and ELA. Within the two-year time frame for students initially in grades 9-11, some may have taken the test before this window, and others may not yet have advanced far enough to have taken these tests within two years. Because we lack full transcripts we cannot know for certain what fraction of the missing test scores are true attrition and what fraction arises simply because the students are not yet far enough along in their studies to have taken these capstone standardized tests, or because they took the tests before our data begins. The above table treats all of these as missing.

As mentioned above, our outcomes are cumulative year 2 outcomes, and while we have measures of the AP outcomes for 100% of students, some of these students had data only for year 1. While we have a measure of AP outcomes for these students in year 1, if we had year 2 of data available their AP outcomes could only have improved. Table D.2-2 re-creates Table

D.2-1 but now the percentage with AP outcomes is reduced to account for those not in the data in year 2.

Table D.2-2 Re-Creation of Table D.2-1 on Availability of Cumulative AP Outcomes by Treatment and Control Group After Considering Students with Only Year 1 Data as Missing

Sample	Treatment	Control	Total	Differential Attrition (T-C)
Full Year 1 Sample #:	6137	6182	12319	
All AP Outcomes from Table D.2-1	90.4	89.3	89.8	1.1

The What Works Clearinghouse (2022, pp. 42-44 and 150-151) designates a study as low-attrition based on overall attrition and a range of acceptable levels of differential attrition. For the above outcomes, with 10.2% overall attrition, a range of acceptable values for differential attrition is any value up to 6.3-10.9%. The 1.1% differential attrition for all of the AP outcomes is well below this range, thus satisfying the low-attrition definition.[2]

Turning to the test score results in Table D.2-1, for math scores that had overall attrition of 53%, the range of differential attrition that meets the low-attrition standards is 0.6-3.0%. The actual differential attrition rate was 4.3%, so the math results do not satisfy the low-attrition requirement. For ELA test scores, the overall attrition rate was 45%, in which case the range of differential attrition that meets the low-attrition standards is 1.8-6.7%. The actual differential attrition rate was 2.3%, so the ELA results satisfy the low-attrition requirement.[3]

In conclusion all the main outcomes related to AP outcomes meet the low-attrition requirement set out by the WWC. For the exploratory outcomes related to test scores, the ELA result but not the math result satisfies the optimistic bound for the low-attrition requirement.

Our pre-registration plan listed an analysis of the outcomes at the end of one year of the CsforEL program as a robustness check. For all of the AP outcomes, we have data on these outcomes for 100% of the sample, with no attrition or differential attrition, meeting WWC requirements for a low-attrition study. This is not the case for test scores in math and ELA. The bottom panel shows results for the math and ELA Grade 11/End-of-Course test scores. Here, the relevant sample excludes students in states with a Grade 11 exam where students in year 1 were in grades 9 or 10. Attrition for ELA scores is 54%, and the WWC low-attrition requirement in this case is differential attrition equal to or below 0.2-2.3%, which is below the actual 4.7%

[2] It is not entirely clear from What Works Clearinghouse (2022) whether for cluster-based RCTs the definition of “overall attrition” is the level of individual attrition or the sum of cluster attrition and individual attrition. For the AP outcomes in this study, that latter sum of attrition is 11% + 10%=21% attrition. At this level of overall attrition, the WWC range for a low-attrition study is differential attrition of 5.3-9.9%. Again, the 1.1% attrition for our AP outcomes easily meets this definition of low-attrition.
[3] If we treated total attrition as the sum of cluster attrition (11%) plus test score attrition, then for ELA test scores, total attrition would be 56%. At this level of overall attrition, the WWC requirements for low attrition are that differential attrition must be in the range of 0.2-2.3%. The ELA attrition, of 2.3%, meets the optimistic bound but not the pessimistic bound.

difference in attrition. For math scores, with overall attrition of 66%, the WWC criteria for low attrition are not met.

Table D.2-3 Percentage Availability of Year 1 Outcomes by Treatment and Control Group

Sample	Treatment	Control	Total	Differential Attrition (T-C)
Full Year 1 Sample (#):	6137	6182	12319	
Percentages Available:				
Enrolled AP CSP	100.0	100.0	100.0	0.0
Passed AP CSP	100.0	100.0	100.0	0.0
Took AP CSP Exam	100.0	100.0	100.0	0.0
Scored 3-5 on AP Exam	100.0	100.0	100.0	0.0
AP CSP Test Score	100.0	100.0	100.0	0.0
AP CSP Grade	100.0	100.0	100.0	0.0
Full Year 1 Sample (#) for Test Scores:	3224	3858	7082	
Percentages Available:				
Grade 11/End-of-Course Math Z-Score	37.1	31.7	34.2	5.4
Grade 11/End of Course ELA Z-Score	48.3	43.6	45.7	4.7

Note: In the bottom panel results for Math and ELA test scores are for restricted samples that exclude students in grade 9 in year 1 of the program in states with a Grade 11 test. They are excluded because these students would not have reached grade 11 during the first year in which the program operated.

References

Cerbin & Kopp, 2006. *International Journal of Teaching and Learning in Higher Education 2006, Volume 18, Number 3, 250-257. http://www.isetl.org/ijtlhe/*

Fernandez, C. (2002). Learning from Japanese Approaches to Professional Development: The Case of Lesson Study: The Case of Lesson Study. *Journal of Teacher Education*, *53*(5), 393-405. https://doi.org/10.1177/002248702237394 (Original work published 2002)

Rock, T. C., & Wilson, C. (2005). Improving Teaching through Lesson Study. Teacher Education Quarterly, 32, 77-92.

Tolbert,S., Stoddart, T., Lyon, EG, & Solís, J. (2014). The next generation science standards, common core state standards, and English learners: Using the SSTELLA framework to prepare secondary science teachers. Issues in Teacher Education 23 (1), 65-90

What Works Clearinghouse. (2022). *What Works Clearinghouse Procedures and Standards Handbook, Version 5.0*, U.S Department of Education, Institute of Education Sciences, Natoina Center for Education Evaluation and Regional Assistance (NCEE)