Annenberg Institute
BROWN UNIVERSITY

EdWorkingPaper No. 26-1540

Unpacking Treatment Effect Heterogeneity in a Large-Scale Dropout-Prevention Experiment

Thomas S. Dee
Stanford University

Chloe R. Gibbs
University of Notre Dame

Rebecca Margolit-Chan
Cornell University

Lillian Morrell
University of Notre Dame

James Sullivan
University of Notre Dame

Timely high school completion is a critical step in successful transitions to post-secondary schooling and the labor market. This randomized controlled trial, deployed in two urban, disadvantaged school districts, tests the effectiveness of a mature and comprehensive student-support program on an unusually large scale (n=2,528). While overall effects on graduation rates are small and statistically insignificant, exploratory analyses reveal notable variation by site and over time, mirroring heterogeneity in treatment timing and program implementation. These findings highlight the challenges of scaling interventions and raise concerns about generalizing results from small-scale, single-site evaluations.

VERSION: August 2026

Suggested citation: Dee, Thomas S., Chloe R. Gibbs, Rebecca Margolit-Chan, Lillian Morrell, and James Sullivan. (2026). Unpacking Treatment Effect Heterogeneity in a Large-Scale Dropout-Prevention Experiment. (EdWorkingPaper: 26-1540). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/zny2-0x48

Unpacking Treatment Effect Heterogeneity in a Large-Scale Dropout-Prevention Experiment

Thomas Dee, Chloe R. Gibbs, Rebecca Margolit-Chan, Lillian Morrell, and James Sullivan *

July 2026

Abstract

Timely high school completion is a critical step in successful transitions to post-secondary schooling and the labor market. This randomized controlled trial, deployed in two urban, disadvantaged school districts, tests the effectiveness of a mature and comprehensive student-support program on an unusually large scale (n=2,528). While overall effects on graduation rates are small and statistically insignificant, exploratory analyses reveal notable variation by site and over time, mirroring heterogeneity in treatment timing and program implementation. These findings highlight the challenges of scaling interventions and raise concerns about generalizing results from small-scale, single-site evaluations.

*Dee: Stanford University & NBER (email: tdee@stanford.edu); Gibbs: University of Notre Dame, W.E. Upjohn Institute for Employment Research & NBER (email: chloe.gibbs@nd.edu); Margolit-Chan: Cornell University (email: rpm248@cornell.edu); Morrell: Wilson Sheehan Lab for Economic Opportunities (email: lmorrell@nd.edu); Sullivan: University of Notre Dame (email: jsulliv4@nd.edu). This research was supported by the University of Notre Dame's Wilson Sheehan Lab for Economic Opportunities (LEO) and MIT's Abdul Latif Jameel Poverty Action Lab (J-PAL). We are grateful to the program provider for their research partnership and to the two participating school districts and the state department of education for their cooperation and assistance with data collection. We also appreciate excellent project management and research assistance from Brendan Perry, Grace Ortuzar, Emily Merola, Anthony Tatarka, Ethan Vonhedemann, and Lucille Broun. Any remaining errors are the responsibility of the authors.

I Introduction

Dropping out of high school is often characterized as a form of "economic death sentence" (Cullen et al., 2013) as the shift towards a knowledge-based economy has made low-education jobs scarcer and less financially viable. This concern reflects a long-standing body of causal evidence, which shows that obtaining a high school diploma confers substantial improvements in lifetime earnings, employment, and health (Card, 1999). In recognition of its importance, high school completion, along with college and career readiness, is now embedded as a key performance indicator in the state accountability systems developed under the Every Student Succeeds Act (ESSA), the 2015 reauthorization of the Elementary and Secondary Education Act.

Although high school graduation rates have increased dramatically in the United States over the last several decades, concentrations of non-completion exist, especially in under-resourced schools and among disadvantaged students. These pockets of persistently low graduation rates have motivated policy-focused research on programs designed to support high school completion among educationally vulnerable students. Structured reviews of this literature sponsored by the U.S. Department of Education's Institute of Education Sciences (IES) underscore the promise of comprehensive programs that feature academic and social-emotional supports, adult mentors who provide individualized attention to at-risk students, and connections between schoolwork and post-secondary college and career opportunities (Dynarski et al., 2008; Rumberger et al., 2017).

In particular, initiatives that integrate such "wraparound" supports and services have attracted considerable attention. For example, the Check and Connect program features adult mentors who monitor student attendance and performance and connect students to aligned social services and academic supports. Early random-assignment evaluations (Sinclair et al., 1998, 2005) found that this approach was effective in reducing dropout when fielded among small populations of students with special learning needs (n=94 and n=147, respectively). However, field-experimental studies situated in somewhat larger and more general student populations have yielded at best mixed results. For example, Maynard et al. (2014) report gains in academic grades, though not attendance, among middle and high school students (n=260). In a study of 9th graders (n=553), Heppen et al. (2018) find that these services had no detectable effects on high school completion or other measures of student engagement and progress. Relatedly, experimental results from a large-scale implementation of Check and Connect among elementary and middle-school students in Chicago Public Schools (i.e., n=9,489) found no effects on multiple indicators of academic success (e.g., test scores and grades), but did find attendance gains limited to the middle school level (Guryan et al., 2020). A similar pattern appears in other large-scale dropout prevention studies. For example, a 4,000-school randomized control trial in Guatemala found only modest short-run reductions in dropout when supports were implemented nationally using existing government systems (Adelman et al., 2025).

This pattern of results is consistent with an established inverse relationship between impact and scale in the broader program-evaluation literature. For example, Kraft (2020) examined 1,942 impact estimates reported in randomized controlled trials of educational interventions and found that the average effect sizes (ES) reported in the smallest studies (i.e., $\leq$ 100 students, ES=0.30; 101-200 students, ES=0.15) are *3 to 6 times* larger than in the largest studies (i.e., $\geq 2,000$ students, ES=0.05). This stark pattern may reflect a selection bias through which higher-impact but expensive interventions are more commonly evaluated in smaller-scale trials. However, these comparative results can also be understood as an indicator of the considerable and well-known difficulty of maintaining a program's implementation fidelity at scale.

This study contributes to our understanding of these issues through the large-scale experimental evaluation of a mature program designed to support the success of educationally vulnerable students through comprehensive high school supports (CHSS). This preregistered randomized controlled trial (RCT) enrolled 2,528 9th-grade students across three district-year cohorts situated in two large urban school districts. The CHSS intervention has multiple features, including mentoring and case management provided by professional staff, coordination of various academic supports (e.g., tutoring and homework assistance), family engagement services, employment opportunities, and assistance with college and career readiness (e.g., support for post-secondary access, resume development, and interview preparation).

Overall, we find that random assignment to this mature program had small and statistically insignificant effects on the primary outcome of interest, high school graduation. However, the unusual scale of this RCT and the richness of the available data allow us to explore key aspects of the program implementation and any underlying heterogeneity in impact. We find that the overall null finding masks extraordinary heterogeneity in program implementation and impact across district-year cohorts. Specifically, in one of the two participating districts, assignment to the program had a large and positive, though only weakly statistically significant, impact on high school graduation for one cohort, followed by a negative impact for a subsequent cohort entering just one year later.

We examine these findings within a well-known framework that characterizes treatment heterogeneity with respect to the "Three C's": client characteristics, program context, and the treatment contrast (Weiss et al., 2014). While student characteristics appear largely consistent across the three site-years, we observe some notable contextual differences. Two of the three district-year cohorts were completing their intended senior year during the first full school year of the COVID-19 pandemic, 2020–2021, while the remaining cohort was in its third year. During this period, the state implemented temporary modifications to graduation requirements, such as changes to assessment policies and diploma criteria.

However, the difference in context resulting from COVID-19 is not a sufficient explanation for the treatment effect heterogeneity we find because we start to see noticeable differences in outcomes across cohorts *before* the onset of the COVID-19 pandemic. In fact, the most stark differences involved the treatment contrasts unique to each district-year cohort, many of which also occurred before the COVID-19 pandemic. Specifically, we find substantial differences in the timing, composition, and intensity of the treatment services students received. In the district-year with positive impacts, program take-up occurred immediately when students began their ninth-grade year, and data shows higher participation in tutoring, academic enrichment, and family engagement relative to other cohorts.

We view this study as making several distinct contributions. Most obviously, this study contributes to the comparatively large experimental literature on the design and impact of dropout-prevention programs targeted to high school students. Notably, this study provides evidence on the impact of comprehensive supports when evaluated on a uniquely large scale. The null results suggest that the enthusiasm for providing high school students with a multifaceted set of supports on a large scale is misplaced if the challenges of consistently implementing these supports with fidelity are not addressed. They also underscore the need for a better understanding of why effective interventions are often difficult to replicate and scale. Alternatively, to the extent the high-fidelity implementation of wraparound services cannot feasibly scale, these results argue for the policy appeal of other program designs. These include widely implemented, low-cost interventions (e.g., text messaging about attendance) as well as more intensive models such as early college access (Berger et al., 2013; Edmunds et al., 2017), cognitive behavioral therapy (Heller et al., 2013, 2017), and career academies (Kemple, 2001). Our findings help inform ongoing debates about the relative effectiveness and scalability of these different approaches.

This study also makes two broader contributions to the program-evaluation literature. First, this study has implications for how we understand the general evidence of an inverse relationship observed between program impact and scale (e.g., List, 2022). In particular, because we find year-to-year variation in implementation and impact within a *single* district-wide site, our results suggest the salience of factors other than positive selection biases in early program-evaluation sites (Allcott, 2015) and the increasing inelasticity of high-quality program inputs at scale (Davis et al., 2017). The evidence presented in this study is consistent with the view that this effect-size/scale relationship reflects the difficulty of consistently maintaining the quality of more complex programs on a larger scale across contexts and time. The broader experimental literature on wraparound services for dropout prevention also supports this interpretation in that the most encouraging evidence of impact appears in the smallest studies.

Second, we view our results as a cautionary tale about the risks of using evaluation evidence in an overly simplistic or credulous manner. For example, the study results presented here illustrate how focusing on a single site in a single year could yield highly divergent conclusions about a program's impact relative to another similar site or even to the same site one year later. This issue is especially relevant given that long-lived research consensuses about particular interventions are often built on the results of relatively few RCTs that rely on data from highly selected sites, small samples, and few time periods. Prior randomized evaluations of Check & Connect, Early College, and Career Academies likewise reveal substantial heterogeneity in both ITT and TOT estimates across programs and contexts, including differences in sign and magnitudes across sites and cohorts. Relatedly, these results also constitute a compelling reminder of the importance of using evaluation evidence not just summatively, but formatively. Specifically, this study's evidence of treatment heterogeneity related to implementation fidelity underscores how rigorous, localized evaluation can be a powerful tool in support of the "continuous improvement" of conceptually compelling programs (Dee, 2024).

II The Intervention

We study the impact of a comprehensive, school-based support program designed to improve high school graduation rates among students at elevated risk of dropout, which we refer to as Comprehensive High School Supports (CHSS). The program was launched in the 1980s and has since grown into a widely deployed and nationally recognized model, currently serving over 4,000 students each year across multiple school districts and states.[1] Participants are typically enrolled at the start of 9th grade and remain engaged through graduation.

At the core of the CHSS model is a full-time, professional case manager assigned to each student. These staff members deliver long-term, individualized mentoring and case management, connecting students to a coordinated web of academic, social, and emotional supports. The CHSS model is guided by an evidence-informed approach that emphasizes relationships, consistency, and responsiveness to student needs.

CHSS services span five key domains:

1. Mentoring and Case Management: Regular one-on-one meetings with case managers, who support academic progress, behavioral goals, and personal development.

2. Academic Support: Tutoring, homework help, and test preparation, particularly for state-mandated graduation exams.[2]

3. Family Engagement: Outreach to caregivers through monthly check-ins, connecting families to community resources, and ensuring continuity of support at home.

4. After-School Enrichment: Activities designed to build life skills, foster positive peer relationships, and promote school engagement.

5. College and Career Readiness: Exposure to post-secondary education pathways and job readiness training, including resume development, interview preparation, and paid internships.

These supports are embedded within a broader model, which integrates school, family, and community partnerships to provide a coordinated system of care for each student. Case managers coordinate closely with teachers, school counselors, and external service providers to align interventions and address both academic and non-academic barriers to success.

During the COVID-19 pandemic, which started in the spring of junior year for the 2017 cohort and the spring of sophomore year for the 2018 cohort, CHSS adapted core services

[1] We do not name the districts or the state due to the provider's preference for anonymity.

[2] In District A, tutoring was provided primarily by phone, while in District B it was delivered in person.

for remote delivery. Mentoring and academic support were provided virtually, and the frequency of remote contact with students and families significantly increased. The program also shifted some of its work toward a stabilizing role by distributing learning materials, technology access resources, and basic necessities such as food, as opportunities for group programming, afterschool activities, and employment, college, and career readiness supports in the community were reduced.

The program operates in districts that primarily serve socioeconomically disadvantaged youth. These districts collectively enroll approximately 40,000 students and invest around $22,000 per student annually.[3] Despite this investment, the average graduation rate in the two districts in 2021 was 70 percent, well below the national average of 86 percent (National Center for Education Statistics, 2025).

Understanding the context of the two districts in the study further underscores the need for targeted intervention. The first district, which we will refer to as "District A", serves over 20,000 students, with half of the families living below the poverty line, more than 60 percent receiving Supplemental Nutrition Assistance Program (SNAP) benefits, and nearly 70 percent of students from single-mother households. Over 10 percent of students have documented disabilities. The other district, or "District B", serves a similarly sized population with comparable demographic and economic profiles.[4] These conditions underscore the relevance of CHSS as an intervention tailored to the multifaceted barriers students face in completing high school.

III Study Design and Enrollment

To determine the effect of this comprehensive program of student and family supports on academic progress and graduation, we conducted a randomized controlled trial (RCT) that included ninth-grade students entering high school from two public school districts, referred to as "District A" and "District B." Both districts are located in urban areas and serve many low-income students who are at high risk of not completing high school. For District A, we enrolled one cohort during the 2017–2018 school year, while for District B, we enrolled two cohorts: 2017–2018 and 2018–2019.

[3]Descriptive statistics come from publicly available district data.

[4]Student demographic data are drawn from the NCES Common Core of Data.

3.1 Constructing the Study Sample and Implementing Randomization

To construct our study sample, we relied on student-level administrative data from the two school districts. At the beginning of the school years during which we were enrolling students, the districts provided the program with lists of incoming ninth-grade students at the 16 schools that offered the intervention. The program applied the pre-specified eligibility criteria and shared de-identified lists with the research team for randomization. These lists included all ninth graders who met at least two out of six predetermined risk factors associated with dropout rates and had accrued no more than 17 days of suspension. The suspension criterion is enforced because extensive suspension days are correlated with unresolved behavioral issues that the program may not be equipped to address. The six factors include:

1. Low socioeconomic status, defined as eligibility for free or reduced-price lunch;
2. Poor school attendance, defined as missing approximately 20-50 instructional days during the prior academic year;
3. Over-age for grade, signifying being behind by 1 or 2 grades;
4. Failure in two or more core academic courses;
5. Subpar standardized test performance, defined as scoring below state proficiency benchmarks or in the lower portion of the distribution on a nationally normed assessment;
6. Having multiple school suspensions.

These eligibility criteria were designed to target students who would be the most likely to benefit from this unique intervention: those who are capable of success but in danger of dropping out of school.

The number of students satisfying these eligibility criteria included 1,005 students from District A, 739 from District B in 2017–2018, and 784 from District B in 2018–2019, totaling 2,528 students. The program could serve about 250 incoming ninth-grade students in District A and 200 in District B in any given year, with the number of available slots varying by school. To determine which students from the pool of eligible students to invite to participate in the program, the research team constructed randomly ordered lists of students stratified by school, cohort, gender, and "risk factor" category (either two risk factors or more than two). The research team sent these randomly ordered lists to the program's senior administrators, who linked the de-identified lists to personal identifying information. Program staff at each school then received the names of students to recruit for program participation, provided in the order generated by the research team. Program staff would reach out to these students in the order that names were received (student names were provided in small batches of about five or so). When students could not be contacted after several attempts or students chose not to participate, program staff would move on to the next student on the list, continuing this process until capacity was reached. Students from the lists who were not invited to participate in the program continued to have access to the other usual services offered by the school district. Each district provided a range of other support services, including school counseling, access to social workers and psychologists, academic tutoring, extended learning time, and referrals to community-based resources. Students could also participate in social-emotional learning (SEL) activities and career and technical education (CTE) programs, which provided industry-aligned training and skill development. However, to our knowledge, no program providing coordinated, individualized, and sustained support – comparable to what students receive with CHSS engagement – was available to students in the control group.

3.1.1 Assigning Students to the Treatment Group: Initial Offer vs. Ever Offer

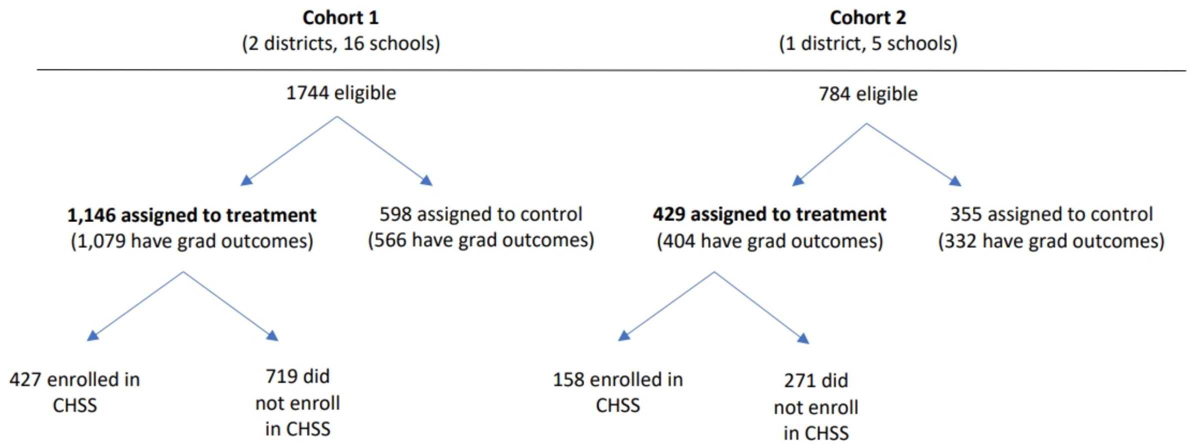
Because of the endogeneity of program takeup in a randomized waitlist setting (de Chaisemartin and Behaghel, 2020), we consider two definitions of assignment to treatment offer. The first approach, "ever offer," assigns the treatment offer to any student whose name was released to the program staff for recruitment, with unreleased names (i.e., names further down on the waitlist) constituting the control group. The second approach, "initial offer," assigns 60 percent of each school's randomly ordered list, in numeric order, to the treatment group and the remaining 40 percent to the control group. This 60 percent-40 percent breakdown was based on projections of program takeup based on information from program staff, prior to distribution of the ordered lists. In this latter approach, some students are assigned to the treatment group even though they were never offered treatment. This approach avoids the bias induced by the endogenous stopping rule because assignment is fixed ex ante. It does, however, introduce administrative non-compliance, as some students classified as receiving an offer treatment were never actually offered the program. This discrepancy is a noncompliance issue, rather than a concern about the assignment variable itself.

Our preferred definition of treatment assignment employs the "ever offer" categorization, which more closely reflects the program's actual process. Following de Chaisemartin and Behaghel (2020), we address the endogenous stopping rule in the "ever offer" approach by dropping the last complier from each waitlist, defined at the school-by-cohort level. Because waitlists close once program capacity is reached, the final offer of each waitlist is extended, by construction, to a student who accepts the offer, creating a mechanical correlation between treatment offers and compliance that can bias conventional ITT estimates. Because there are 21 school-by-cohort waitlists, implementing this adjustment reduces our full sample from 2,528 to 2,507 students. Of these 2,507 students, we observe graduation status is observed for 2,381. After excluding the 21 last compliers from each waitlist who have observed graduation outcomes, the final analytic sample is 2,360. In our robustness section, we also present results using the "initial offer" definition and show that our findings are qualitatively similar, though less precise, under this alternative approach.

Figure 1 displays the number of program-eligible students in each cohort and their assignment to treatment or control groups based on the "ever offer" approach. A total of 2,528 students were eligible for the program based on eighth-grade measures, 1,744 from the 2017 cohort and 784 from the 2018 cohort. Of these, 1,575 were offered treatment, and 953 are in the control group. However, our primary analysis that examines the effect of the program on high school graduation relies on a large subset of this study sample, because graduation status is missing in school district records for 5.8 percent of the students. We examined whether treatment status was related to the likelihood that a student's graduation status was missing. In general, we do not find noticeable differences in missingness by treatment status (see appendix Tables A.1 and A.2 for additional detail.).[5] We observe graduation status for 1,645 students from the 2017 cohort and 736 from the 2018 cohort. Therefore, our analysis sample includes 2,381 students with observed graduation status prior to the last-complier exclusion described above, yielding a final analytic sample of 2,360 students.

[5]Specifically, we define missingness as having transferred outside of the state or having an unknown status. Students with unknown statuses are students for whom the district did not receive any documentation. While we find little difference in missingness by treatment status, the results in Appendix Table A.2 do show that students who were assigned to the treatment group in the second cohort of District B were approximately 3.3 percentage points (roughly 73 percent) more likely to transfer to a high school equivalency program compared to those in the control group. However, we find that this effect is driven by those who did not take up the program.

Figure 1: Study Sample by Cohort, District, Treatment Status, and Program Take-up



Notes: This figure displays the flow of students through the study by cohort, district, treatment status, and CHSS program enrollment. Treatment assignment uses the "ever offer" definition, which classifies a student as treated if their name was ever released to program staff for recruitment. Parentheses reflect the subset of students within each group for whom high school graduation outcomes are observed in administrative records. Appendix Table A.1 reports regression-adjusted differences by treatment status in graduation-outcome missingness, reasons for missingness, and match to National Student Clearinghouse records. Inclusion in the analytic sample does not differ significantly by treatment status.

3.1.2 Program Takeup

As shown near the bottom of Figure 1, many students assigned to receive the program did not enroll (about two-thirds). For instance, in the 2017 cohort, 1,146 students were assigned to the treatment group, but only 427 ultimately enrolled in the program, as indicated by having a program start date.[6] Appendix Table A.3 presents regression-adjusted estimates of the effect of treatment assignment on program enrollment. Across all specifications, the first stage is large in magnitude and statistically significant, with assignment to treatment increasing the likelihood of enrollment by approximately 34-40 percentage points depending on district and cohort.

[6]Students are classified as having enrolled in the program if they are recorded in the data as having a program start date.

IV Data and Methods

4.1 Characteristics of the Study Sample

Our empirical analysis leverages student-level administrative records from the two public school districts where the study was implemented. Before randomization, the school districts provided the program with files containing baseline characteristics of entering ninth-grade students, which were subsequently shared in de-identified form with the research team.

The administrative data provided by the school districts include key demographic indicators such as birth date, gender, race, ethnicity, language spoken at home, eligibility for free and reduced-price lunch, English language learner status, and the total number of risk factors. In District A, low socioeconomic status is defined as household income below the federal poverty line, while in District B, it is indicated by eligibility for free or reduced-price lunch. District B also provides information on whether a student has a documented disability, defined as having either a 504 Plan or an Individualized Education Program.

Table 1 reports the baseline characteristics for the 2,507 students in the full study sample and the 2,360 students in our analysis sample with observed graduation outcomes, both after excluding the last complier in each waitlist as described above. In the first three columns, we report means for the full sample, the control group, and the treatment group, respectively. In the fourth column, we report the regression-adjusted difference between these means, including as covariates: whether the student is female and whether the student had over two risk factors, as well as school at baseline and cohort fixed effects.

The results show that these characteristics are balanced across the two groups–none of the differences are statistically significant, either for the full sample or for the analytic sample. These baseline characteristics also emphasize how the CHSS program targets extremely disadvantaged students. Nearly all of the students live in poor households; half have more than two risk factors (only two are required for eligibility); for a quarter of them, English is not their native language; and the predicted graduation rate (based on observable baseline characteristics) is 73 percent, well below the national average.

4.2 Outcomes

Our primary outcomes, specified in advance in our longer-term outcome pre-analysis plan[7], are two measures of high school completion — indicators for graduation within either four or five years. Graduation is determined by receipt of a diploma by June 30 of the relevant year. These outcomes are constructed using longitudinal student-level data from the participating school districts, supplemented with statewide administrative records on high school enrollment and exit. If a student is identified as having graduated in either district or state data, we classify them as having completed high school. If the district data are missing or incomplete, we rely on state records to determine graduation status.[8]

Additionally, we examine several closely related outcomes that shed light on students' academic progress and pathways through high school. We consider several different indicators of whether a student is on-track to graduate that are based on a review of related literature (Allensworth and Easton, 2005; Kemple and Stephenson, 2013). Our measures of being on track are functions of several indicators of school progress, including whether, in their first year, a student drops out, earns fewer than 2.5 credits, is chronically absent, or passes at least one graduation exam.

We linked these student records to data from the National Student Clearinghouse (NSC), enabling us to observe post-secondary education outcomes, such as enrollment in two- or four-year colleges or vocational programs, the name of the institution, the number of semesters enrolled, whether enrollment was full or part-time, and graduation status[9].

Throughout, we distinguish between confirmatory analyses specified in advance in our pre-analysis plans and exploratory analyses that were not. Our confirmatory analyses comprise

[7] There are two pre-registration plans associated with this study. The first registration covers shorter-term (i.e., in high school) outcomes, focused on a measure of being "on track" to graduate, and is registered under AEARCTR-0007124. The second registration covers longer-term outcomes and is registered under AEARCTR-0002384. The second plan was posted to the registry prior to the receipt of any high school graduation data.

[8] Students are classified as having graduated only if they earned a diploma recognized by the state as fulfilling high school graduation requirements. Students who earned a High School Equivalency (HSE) certificate or a Commencement Credential are not considered graduates. Students whose final status indicates dropout, long-term absence, aging out without a diploma, or withdrawal without documentation are also classified as not graduated. Students recorded as transferring out of the district are considered graduates only if diploma receipt is observed in state data.

[9] The National Student Clearinghouse is a nonprofit organization that works with more than 3,600 post-secondary institutions to provide data on student enrollment and degree completion. The NSC-participating colleges enroll 98 percent of all post-secondary students in the US (National Student Clearinghouse, 2025).

the intent-to-treat effects of program assignment on our primary outcomes, high school graduation and the on-track-to-graduation measures, along with program take-up, and pre-registered subgroup analyses of these outcomes along five dimensions: gender, race and ethnicity, age, baseline risk-factor category, and predicted likelihood of the outcome. All other analyses are exploratory. These include the individual components of the on-track measures, the analyses disaggregated by district and cohort, the heterogeneity analysis by disability status, the examination of program implementation and treatment contrast, and the post-secondary outcomes.

4.3 Provider Data on Participation and Activity Engagement

The implementing partner provided detailed administrative records on student participation in the CHSS program. These records included information on the duration of each student's engagement, the frequency of contact with their assigned case manager, and logs of participation in specific program activities. For each activity, the data captured its type, frequency, and duration. The provider also shared staff-level data, including unique identifiers for case managers and indicators of staff turnover during the study period.

4.4 Methods

Given the random assignment of students to treatment and control groups, we estimate the program's impact using an intent-to-treat (ITT) model. In particular, we estimate:

$$Y_{ics} = \beta_0 + T_{ics}\beta_1 + X_{ics}\beta_2 + \theta_c + \lambda_s + \epsilon_{ics} \tag{1}$$

Where Y_{ics} is an indicator for one of our outcome measures for student i. T_{ics} is an indicator for whether student i from the randomly ordered waitlist in cohort c and school s was ever offered to enroll in the CHSS program.[10] The vector X_{ics} includes a set of person-level characteristics collected at baseline, including gender, race/ethnicity (White, Black, or Hispanic), whether English is a second language, poverty status, indicators for age at

[10] As noted earlier, for all of our main analyses, we define treatment status based on the "ever offer" categorization, but in our robustness section, we present results using the "initial offer" definition, showing that the choice of treatment definition does not substantively affect our results.

enrollment (15 or 16+), and whether they have more than two risk factors.[11] θ_c and λ_s are cohort and school fixed effects respectively, and ϵ_{ics} is an error term. β_1 measures the average effect of the program for students who are offered the opportunity to enroll. All estimates are based on the analytic sample of 2,360 students described in Section 4.1, which excludes the last complier in each waitlist following de Chaisemartin and Behaghel (2020). Our preferred specification uses sandwich standard errors clustered at the school x gender x risk factor x cohort x district level following de Chaisemartin and Behaghel (2020). We also report heteroskedasticity-robust standard errors for completeness.

To determine the effect of the program on those who actually participated (treatment-on-the-treated, TOT), we estimate the effect of program participation using an instrumental variables (IV) model, where treatment assignment serves as an instrument for program participation. The TOT model is specified as:

$$Y_{ics} = \gamma_0 + \hat{P}_{ics}\gamma_1 + X_{ics}\gamma_2 + \theta_c + \lambda_s + \eta_{ics} \tag{2}$$

where $\hat{P}_{ics}$ is an indicator for whether student i completed the program enrollment process. Because enrollment is potentially endogenous, we instrument P_{ics} using treatment assignment T_{ics} The corresponding first-stage equation is:

$$P_{ics} = \delta_0 + T_{ics}\delta_1 + X_{ics}\delta_2 + \theta_c + \lambda_s + \nu_{ics} \tag{3}$$

The 2SLS estimator employs the predicted values from the first stage to recover γ_1, the local average treatment effect among compliers. The first-stage relationship between treatment assignment and program enrollment is strong across all three district-cohort samples, with F-statistics well above conventional thresholds (Appendix Table A.3), satisfying the relevance condition for the IV approach.

[11] Some of our observed characteristics gender, race/ethnicity, English as a second language, poverty status, age at enrollment, and having more than two risk factors–have missing data for a small number of observations (2 or less missing observations). Rather than dropping these observations, we include in our specification an indicator for whether the specific characteristic is missing as well as an interaction of this indicator with the value of the variable.

Table 1: Baseline Balance, Full and Analytic Samples

	All	Control	Treatment	Adj. Difference
	Means	Means	Means	Treatment - Control
	(a) Full Study Sample			
White	0.162	0.173	0.155	-0.012
				(0.015)
Black	0.530	0.518	0.538	0.017
				(0.021)
Hispanic	0.206	0.200	0.210	-0.0018
				(0.016)
English Second Language	0.266	0.276	0.259	-0.017
				(0.018)
In Poverty	0.933	0.931	0.935	0.0039
				(0.010)
Enrollment Age (Years)	14.10	14.11	14.09	-0.023
				(0.025)
Predicted Graduation Rate	0.727	0.729	0.727	-0.0016
				(0.0033)
Female	0.475	0.472	0.476	0.0011
				(0.020)
Over Two Risk Factors	0.529	0.529	0.529	0.0037
				(0.020)
Total Obs.	2,507	953	1,554	2,507
	(b) Analytic Sample			
White	0.160	0.175	0.150	-0.018
				(0.016)
Black	0.541	0.530	0.548	0.016
				(0.021)
Hispanic	0.203	0.196	0.207	-0.00024
				(0.017)
English Second Language	0.250	0.262	0.242	-0.020
				(0.018)
In Poverty	0.933	0.930	0.935	0.0036
				(0.011)
Enrollment Age (Years)	14.09	14.11	14.08	-0.031
				(0.026)
Predicted Graduation Rate	0.729	0.730	0.729	-0.00057
				(0.0033)
Female	0.479	0.473	0.482	0.0064
				(0.021)
Over Two Risk Factors	0.529	0.524	0.532	0.013
				(0.021)
Total Obs.	2,360	898	1,462	2,360

Notes: This table reports baseline characteristics for the full study sample and the analytic sample. The difference column reports the regression-adjusted difference in means between the treatment and control groups, where ever-offer is used as the treatment definition and the regression adjusts for whether the student is female and whether the student had over two risk factors. The full sample includes all students who were randomized for the study, excluding the final treated student in each randomized waitlist to account for the stopping rule inherent in randomized waitlist designs (de Chaisemartin and Behaghel, 2020). The analytic sample restricts the full sample to students for whom we observe the main graduation outcomes. Standard errors are in parentheses. *** p<0.01, ** p<0.05, p<0.10. Predicted graduation rate is a predicted value of the graduation rate constructed using the other covariates in the table following the repeated split-sample (RSS) procedure of (Abadie et al., 2018).

V Results

In this section, we report the treatment effects for our primary outcome, high school graduation, for our full analytic sample as well as separately by district and cohort.

5.1 High School Completion

We estimate treatment effects on four- and five-year high school graduation for the full sample and across three separate district-cohort groups: one cohort that entered high school in the 2017–18 school year in District A, and two entering cohorts, in 2017–18 and 2018–19, in District B. We use the "ever offer" definition of treatment and present two approaches to statistical inference. Our preferred specification reports bias-corrected sandwich (HC2) standard errors clustered at the cohort-school-risk factor-gender level. We also report heteroskedasticity-robust standard errors for completeness (Abadie et al., 2023)[12].

The pooled estimates presented in column 1 of Table 2 suggest that the program has no effect on four- or five-year graduation. For our primary outcome, four-year graduation, the point estimate is very small (-0.014) and not statistically different from zero. We can rule out positive effects greater than 2.5 percentage points, or a 3.4 percent increase relative to the mean for the control group. This null finding for the pooled sample, however, masks important variation in the treatment effect across districts and cohorts. For District A (column 2), we also see little evidence of a treatment effect. For District B (columns 3 and 4), the treatment effects differ sharply across the two cohorts. Students in the 2017 cohort who were offered the program were about 5.7 percentage points more likely to graduate, which is a large effect compared to the control group mean of 73 percent, although this effect is only marginally significant using the heteroskedasticity-robust standard errors. In contrast, for the 2018 cohort, treatment group students were significantly less likely to graduate in four (-8.8 percentage points) or five (-7.4 percentage points) years, and the effects for both four- and five-year graduation are statistically significant.

[12]In a small share of specifications (approximately 6 percent of estimates), this clustered HC2 variance estimator cannot be reliably computed because the bias-corrected sandwich standard errors returns a variance matrix that is non-symmetric or highly singular, a numerical limitation that arises in cells with sparse cluster-by-fixed-effect structure. In these cases, we leave the standard error blank and we omit the corresponding whiskers in the ladder-plot figures. As shown in Appendix Table A.4, the standard errors for our main estimates are very similar across a range of alternative inference procedures, suggesting that the choice of standard error is unlikely to be consequential for the affected estimates.

Because fewer than half of the students who were offered the treatment took it up, we also considered the effect of the program for those who participated by estimating treatment-on-the-treated (TOT) effects (Appendix Table A.5). Since the TOT estimates are essentially the ITT estimates scaled up by the take-up rate, it is not surprising that we see similar heterogeneity across cohorts and districts. For the first cohort in District B (column 6), the program led to a 16.2 percentage-point increase (p-value: 0.088) in the likelihood of graduating in four years and a 17.1 percentage-point increase (p-value: 0.069) after five years. While only marginally significant, these estimates suggest that the program effect is quite large for those who participate–the effect on four-year graduation, for example, reflects a 22 percent increase relative to the control complier mean. Likewise, the TOT estimates suggest that the program effect for the second cohort in District B is large and negative, implying that students who enrolled in the program experienced meaningfully worse graduation outcomes than they would have otherwise. We note that the derived control complier mean for District B C18 exceeds 1, which is outside the unit interval for a binary outcome. This reflects the combination of a large negative TOT estimate and a modest first stage for that cohort, which together cause the derived statistic to behave poorly. It should not be interpreted as a graduation probability but is retained in Table A.5 for completeness. We do not attempt to identify causal effects for non-compliers, as the IV framework does not support that inference. However, descriptively, students in the 2018 District B treatment group who did not enroll exhibited particularly unfavorable baseline characteristics relative to comparable students in other cohorts (see Appendix Table A.6 for baseline characteristics of program compliers). This raises the possibility that assignment to treatment may have altered access to alternative supports during a critical period without substituting meaningful program engagement. This interpretation remains speculative and should be treated with caution.

Table 2: ITT Effect of Program on 4- and 5-Year Graduation, Pooled and By District

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(a) 4-Year Graduation			
	(1)	(2)	(3)	(4)
Treatment	**-0.014**	0.012	0.057	-0.088
	(0.020)	(0.028)	(0.035)	(0.032**)
	[0.018]	[0.027]	[0.033*]	[0.032***]
Control Mean	**0.7294**	0.7143	0.7257	0.7470
N	**2,360**	941	688	731
	(b) 5-Year Graduation			
	(1)	(2)	(3)	(4)
Treatment	**-0.006**	0.018	0.060	-0.074
	(0.020)	–	(0.035)	(0.040*)
	[0.017]	[0.027]	[0.033*]	[0.031**]
Control Mean	**0.7572**	0.7508	0.7342	0.7801
N	**2,360**	941	688	731

Notes: Estimates are intention-to-treat (ITT) effects estimated via OLS, with treatment defined using the ever-offer definition. All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, in poverty, age at enrollment, and whether they have more than two risk factors. The analytic sample includes the 2,360 students with observable graduation outcomes that remain after dropping the last complier from each waitlist to account for the endogenous stopping rule in randomized waitlist designs (de Chaisemartin and Behaghel, 2020). Standard errors in parentheses ( ) are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level; those in brackets [ ] are heteroskedasticity-robust standard errors. In a small share of cells, the clustered HC2 standard error is left blank because it cannot be reliably computed, as discussed in footnote 12. The heteroskedasticity-robust standard error in brackets is reported in all cases. *** p<0.01 ** p<0.05 * p<0.10. Bolded analyses were pre-registered.

We conduct several additional analyses to probe the robustness of our main findings. First, we consider how sensitive our results are to an alternative definition of treatment, using "initial offer" rather than the "ever offer" definition of treatment used for our main analyses. The estimates using "initial offer" (Appendix Table A.7) are very similar to the estimates using "ever offer" (Table 2), both in terms of the size and precision of the estimates.

This is not surprising given that the fraction of the sample assigned to treatment using the "initial offer" approach (60 percent) is very close to the fraction assigned to treatment using the "ever offer" approach (62 percent).

We also assess robustness to the method used to address the endogenous stopping rule in the randomized waitlist design. Appendix Table A.8 presents ITT estimates using the Doubly Reweighted Ever-Offer (DREO) estimator proposed by (de Chaisemartin and Behaghel, 2020) as an alternative to our preferred approach of dropping the last complier from each waitlist. The results are qualitatively similar to our main estimates.

In addition, our preferred specification reports bias-corrected sandwich (HC2) standard errors clustered at the cohort-school-risk-factor-gender level. To investigate how sensitive our findings are to our approach to statistical inference, we use the following alternative approaches reported in Appendix Table A.4: heteroskedasticity-robust (unclustered) standard errors, school-level clustering, school-gender-risk factor clustering, bootstrapped unclustered standard errors, bootstrapped standard errors with school-level clustering, and bootstrapped standard errors clustered at the school-gender-risk factor level. As shown in Appendix Table A.4, the standard errors for our main point estimates are very similar across these different approaches.

VI Understanding District and Cohort Heterogeneity

While we find null results for our full, pooled sample, we find evidence suggesting that the program had very different effects across districts and cohorts. In particular, in one of the two participating school districts, assignment to the program had a positive impact on high school graduation in one cohort but a negative impact on a subsequent cohort, entering just one year later.

We explore these differences further within a well-known framework that characterizes treatment heterogeneity with respect to the "Three C's": client characteristics, program context, and the treatment contrast (Weiss et al., 2014). The first of these three C's might help explain our heterogeneous treatment effects if the types of students that are served by the program differ noticeably across districts and/or cohorts, and perhaps the program affects different types of students differently–we did find some evidence of heterogeneous treatment effects by student characteristics (see Section VII). However, the baseline characteristics of the students in the study do not differ noticeably across district and cohort (shown in Appendix Tables A.9 and A.10). The similarity is not surprising given the structured eligibility for the CHSS program, and it suggests that differences in students do not play a major role in explaining the heterogeneous effects we find across districts and cohorts.

There is also reason to be skeptical about the second C, context, as an important explanation for the heterogeneity we find. The districts are geographically adjacent and in the same state. Both districts for this study are large and primarily serve socioeconomically disadvantaged youth. They serve a similarly sized population with comparable demographic and economic profiles. Importantly, there may be meaningful differences in district-level factors that are difficult to observe or quantify in administrative data, such as school leadership, local resource constraints, community-based programs, and the availability or quality of supplemental academic supports.

There is, however, an important difference in context across cohorts in that COVID-19 occurred at different points in their high school experience–COVID hit seven semesters after the start of 9th grade for the 2017 cohort, while it hit five semesters after the start of 9th grade for the 2018 cohort. During this period, the state implemented temporary modifications to graduation requirements, such as changes to assessment policies and diploma criteria, including requirements tied to graduation exams. However, the difference in context resulting from COVID-19 is not a sufficient explanation for the treatment effect heterogeneity we find because we start to see noticeable differences in early outcomes across cohorts *before* the onset of the COVID-19 pandemic, as we show in Section 6.3.

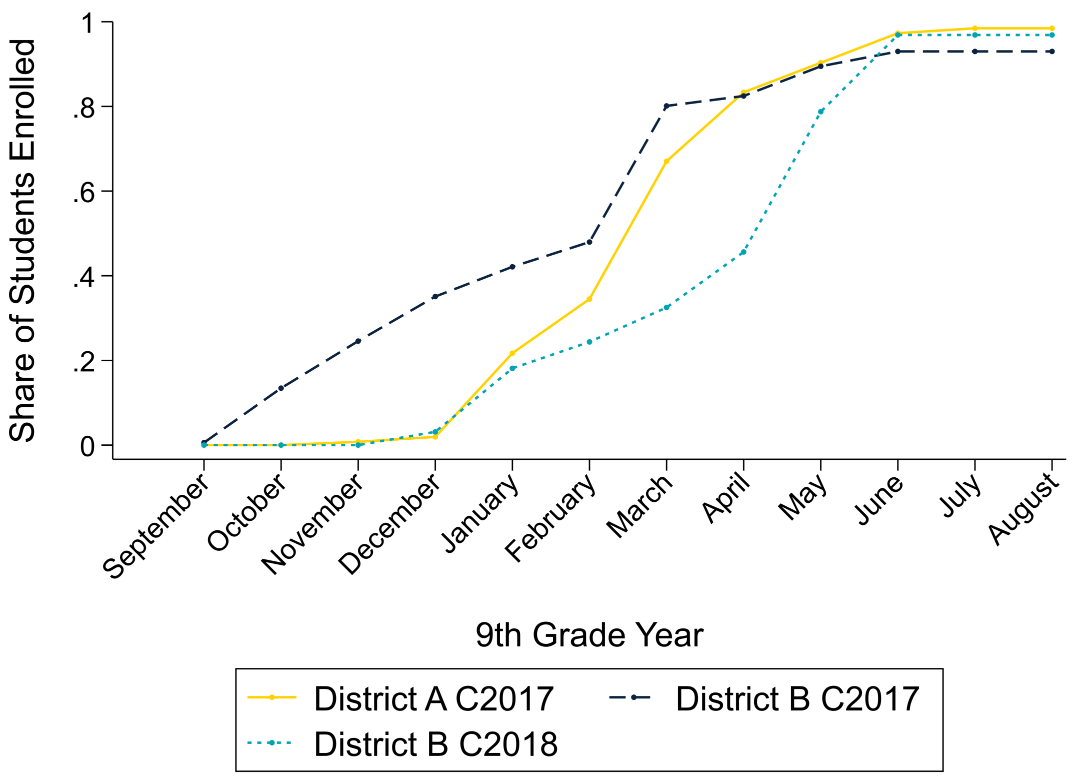
To assess the role that treatment contrast might play in the differences in treatment effects that we find, we consider how the actual treatment differed across cohorts and districts along several dimensions, including the timing of enrollment during the school year and the content and intensity of the treatment.

6.1 Enrollment Timing

One of the more noticeable differences in treatment contrast across districts and cohorts was the timing of enrollment. The goal of the CHSS program was to enroll at-risk students into the program as early on in their high school experience as possible in order to address issues that might prevent students from getting too far off track. That is why eligibility for the program was based on factors prior to enrollment into high school and why the program aimed to identify the most at-risk students and enroll them into the program early on in high school.

As is shown in Figure 2, there is considerable variation in the timing of enrollment across districts and cohorts, with the 2017 cohort in District B enrolling students much earlier. This figure reports the share of all students who eventually enroll in the CHSS program in a particular district and cohort who have enrolled by each month since starting ninth grade. For the 2017 District B cohort, over a third of the students who would enroll in CHSS had already enrolled by December of freshman year, which is in sharp contrast to District A or the 2018 cohort, District B, neither of which had enrolled even five percent of eventual enrollees by December. The slow enrollment for the 2018 cohort of District B is particularly noticeable, as less than half of the eventually enrolled students had enrolled by April of their freshman year.

Figure 2: Timing of Program Enrollment During 9th Grade, by District and Cohort



Notes: This figure reports the cumulative share of students who ever enrolled in the CHSS program, by month since entry into 9th grade and by district and cohort. Program entry is defined using the CHSS program start date recorded by program staff.

There were a number of reasons for the unusually slow enrollment of students into CHSS for the 2018 cohort of District B. Initially, data transfer issues delayed the creation of the lists of eligible students, which in turn delayed outreach by case managers. In addition, the program faced significant staffing and implementation challenges in this cohort. In an effort to ensure students were enrolled during their ninth-grade year, some students were enrolled before a designated case manager was assigned. As a result, many received only limited services early on and experienced delays in accessing core components of the program, such as mentoring and academic support.

Regardless of the reason for the delay, these data suggest that enrollment timing could be an important factor for determining program impact — we see suggestive evidence of positive program impacts for the one district-cohort that enrolled students earlier in their first year of high school, but little evidence of impact, or even a negative impact, in district-cohorts where students were enrolled much later in the academic year.

To further examine the role of outreach timing, Table A.11 uses students' position on the randomized waitlist as a proxy for when the program was likely to have contacted them. Specifically, we define students as being at the 'top of the waitlist' if their assigned position is at or below the position of the last student who enrolled in the program within the first semester, by January of their freshman year. Panel (a) shows that being at the top of the waitlist strongly predicts program take-up in District A and District B Cohort 2017 with F-statistics of 21.76 and 17.39 respectively, well above conventional thresholds for instrument strength. In sharp contrast, the F-statistic for District B Cohort 2018 is essentially zero (0.09), indicating that waitlist position had no meaningful relationship to take-up in that cohort. This finding is evidence of program disruption. In a normally functioning program, students near the top of a randomly ordered list should be the first to be contacted and enrolled, and the breakdown of this relationship in District B Cohort 2018 is itself direct statistical evidence of the implementation disruption described above.

Panel (b) documents that being at the top of the waitlist predicts significantly lower rates of transfer to high school equivalency (HSE) programs among treated students, a 3.1 percentage point reduction in the pooled sample ($p<0.05$).[13] Panel (c) shows that waitlist position is also positively associated with graduation within four years in District B, with

[13] High school equivalency programs offer students an alternative pathway to obtaining a high school credential outside the traditional diploma track. Students who transfer to these programs are not classified as high school graduates in our primary outcome.

point estimates of 0.087 and 0.073 for Cohort 2017 and Cohort 2018, respectively, though these estimates are not statistically distinguishable from zero. While the graduation panel is suggestive rather than conclusive, the pattern across all three panels tells a consistent story: in District A and District B Cohort 2017, where the program worked through its list in an orderly way, being at the top of the list predicted take-up, protected against HSE transfers, and was associated with better graduation outcomes. In District B Cohort 2018, this process broke down, list position no longer predicted who got reached, and students left at the bottom of the list were without support at a critical juncture in their freshman year.[14]

A potential concern with this analysis is that our definition of "top of the waitlist" depends on the position of the last student to enroll by January, which is itself an outcome of the program's recruitment process and therefore endogenously determined. To assess whether this choice drives our results, Appendix Table A.13 reproduces the analysis using a pre-determined definition that classifies students at or below the median waitlist rank within each school-by-cohort waitlist as being at the top of the list. The results are qualitatively similar across both definitions, suggesting that our findings are not an artifact of the endogenous January-enrollment cutoff. The main exception is the first stage for District B Cohort 2017, which is less precisely estimated under the median-rank definition, though the broader pattern across take-up, HSE transfer, and graduation outcomes is preserved.

Enrollment timing also appears to be related to early outcomes such as credit accumulation, chronic absenteeism, and exam-taking. The treatment effects for these early outcomes mirror the graduation results: the cohort that enrolled early and received robust services (District B 2017) shows gains across these early outcomes, whereas delayed or fragmented enrollment (District B 2018) coincides with null or negative outcomes.

6.2 Other Variation in Program Implementation

Using additional program data from the organization that runs CHSS, we explored other differences in the way in which the program was implemented across districts and cohorts.

[14] Appendix Table A.12 reports early academic outcomes separately for students who transferred to HSE programs and those who did not. Across all districts and cohorts, HSE students were substantially more off-track in their first year, earning on average fewer than 1.4 credits compared to over 4.5 credits for non-HSE students, with higher rates of chronic absenteeism and far lower rates of taking or passing graduation exams. This is consistent with the interpretation that students not reached in time by the program in District B Cohort 2018 were already academically vulnerable and, absent program support, pursued HSE as an alternative pathway.

Even though the program has very detailed procedures and manuals, and offers a consistent training program for their case managers, inevitably variation in implementation and local adaptation arises. Appendix Table A.14 reports the fraction of students who enrolled in CHSS that participated in various program activities by district and cohort. We also examined differences in treatment contrast by looking at service engagement on the intensive margin–how many minutes did the students spend engaged in each of these activities (Table 3).

The most obvious difference across these groups is in tutoring. 84 percent of students enrolled in the 2017 cohort in District B participated in tutoring, which is 18 percentage points more than the participation rate for enrolled students in District A and 24 percentage points more than the rate for those in the 2018 cohort of District B. Also, while nearly all students participated in at least some mentoring services, engagement was much greater for the students in the 2017 cohort in District B than for the other students (Panel A of Table 3). Specifically, the 2017 District B students spend 215.3 more minutes (14.4 percent) in mentoring than students in District A, and 405.6 more minutes (31 percent) than those in the 2018 cohort of District B. Similarly, while the program had contacted the family for nearly all students in the program (Appendix Table A.14), the time spent connecting with family was noticeably greater in the 2017 District B cohort than the other district/cohort (Table 3). For all three of these activities–tutoring, mentoring, and family contact–the district-cohort that engaged students the most in these services was also the district with the largest treatment effect.

Table 3: Program Activity Engagement Among Enrollees, by District and Cohort

	District A C17	District B C17	District B C18	Dist. A C17 minus Dist. B C17	Dist. B C18 minus Dist. BC17
	Means			Means (Standard Errors)	
	(a) Minutes, Unconditional on Participation in Activity				
Tutoring Sessions	503.9	1236.3	302.7	-732.4***	-933.6***
				(145.3)	(138.6)
After School Activities	2607.4	1931.4	1775.6	676.0*	-155.8
				(369.8)	(326.1)
Post HS Exploration	108.6	22.16	26.41	86.4***	4.24
				(24.4)	(10.6)
Mentoring Sessions	1491.2	1706.5	1300.9	-215.3**	-405.6***
				(107.8)	(128.1)
After School Activities	201.3	40.12	44.81	161.2***	4.69
				(28.5)	(14.1)
Employment Related Activities	31.30	52.95	28.69	-21.7	-24.3
				(21.5)	(25.9)
Contacting Family	110.0	314.5	178.6	-204.5***	-136.0***
				(21.1)	(23.1)
Student Enrichment	530.8	65.12	93.75	465.7***	28.6
				(76.3)	(41.8)
Other/Unknown	1834.5	1766.1	1606.8	68.4	-159.3
				(338.9)	(310.3)
Total Obs.	258	171	160		
	(b) Minutes, Conditional on Participation in Activity				
Tutoring Sessions	755.8	1468.1	504.5	-712.2***	-963.5***
				(171.1)	(160.8)
After School Activities	2912.2	2752.2	2630.5	159.9	-121.7
				(430.3)	(431.6)
Post HS Exploration	329.6	126.3	222.4	203.3***	96.0
				(70.6)	(64.3)
Mentoring Sessions	1491.2	1747.4	1317.4	-256.1**	-430.0***
				(107.8)	(128.3)
After School Activities	509.2	228.7	217.2	280.5***	-11.4
				(64.0)	(52.7)
Employment Related Activities	118.8	905.5	573.8	-786.8***	-331.7
				(226.8)	(340.7)
Contacting Family	116.8	324.0	180.8	-207.2***	-143.2***
				(21.3)	(23.3)
Student Enrichment	925.3	232.0	441.2	693.4***	209.2
				(139.1)	(170.7)
Other/Unknown	2161.2	2626.1	2448.5	-465.0	-177.6
				(414.9)	(424.0)
Total Obs.	258	171	160		

Notes: The sample includes only students who enrolled in the CHSS program. Panel (a) reports mean minutes spent in each activity unconditional on participation, while Panel (b) reports means conditional on having participated in that activity at least once. Difference columns report regression-adjusted differences in means between district-cohort groups. Standard errors in parentheses are heteroskedasticity-robust. *** p<0.01 ** p<0.05 * p<0.10.

Some activities appear to have been less emphasized for the 2017 District B cohort. For example, students in this group spent much less time on afterschool activities, the category that accounted for the most time for all three groups, than students in District A.

It is important to note that the data on services that students experience are only available for those in the treatment group who took up the program. Thus, we do not have sufficient information to fully characterize the treatment contrast. Nevertheless, we are not aware of any other comprehensive support program of comparable scope or intensity available to students in these districts during this period. The sharp differences we observe in the implementation of CHSS strongly suggest that participating students experienced variation in treatment contrast, characterized by the nature of and intensity of program services, across sites and cohorts.

A related concern is the potential for crossover among control group students into similar interventions, services, and supports. Both districts offered a range of supports to students throughout this time, including guidance counseling, school psychologists, social workers, career and technical education, after-school and extended learning opportunities, and tutoring. In addition, students may have had access to services provided through community-based organizations and partnerships with non-school-based providers. To the extent that some of these services were introduced or expanded later in the study window, control group students in the later cohort may have received meaningfully more support than those in the earlier cohort - support that would not be captured in our treatment contrast. The possibility could attenuate the estimated treatment effect over time.

6.3 Program Impact on Early Outcomes

To better understand how CHSS affects high school completion and why program impacts differ across districts and cohorts, we examine a set of early and intermediate indicators of student progress. Prior research shows that early academic momentum, particularly during ninth grade, is strongly predictive of eventual graduation. Accordingly, we analyze program impacts on credit accumulation, attendance, exam participation and performance, persistence, and composite measures of being on track to graduate. Table 4 summarizes intent-to-treat estimates for these outcomes for the pooled sample and separately by district and cohort.

Early credit accumulation is a key marker of progress toward graduation, given that students must meet a minimum credit threshold to earn a diploma. For the pooled sample, estimated impacts on credits earned in Year 1 are small and statistically indistinguishable from zero. However, there is meaningful heterogeneity across districts and cohorts. In particular, students in the 2017 cohort in District B earn 0.36 (8.3 percent) more credits by the end of their first year of high school, while estimated effects for the District B cohort of 2018 are negative and imprecise. These early differences mirror the heterogeneity observed in longer-run graduation outcomes. We also examine cumulative credit accumulation in years 1 and 2, though these estimates are statistically imprecise.

We next examine chronic absenteeism, defined by the state education department as missing at least 10 percent of enrolled school days in a given academic year. For the pooled sample, the estimated reduction in chronic absenteeism in Year 1 is modest and not statistically significant in District A and District B cohort 2018. However, for District B 2017, the point estimate of -0.058 is statistically significant at the 5 percent level, a 16.2 percent reduction in chronic absenteeism from the control group mean. [15] Estimated effects attenuate in later years and are less precisely measured, in part due to missing attendance data in subsequent academic years.[16]

We also assess impacts on participation in and passage of required graduation exams, which are an important early benchmark for high school completion in this state. For the pooled sample, estimated effects on both taking and passing at least one exam are small and imprecise. In contrast, students in the 2017 District B cohort are approximately 9.8 percentage points (13.1 percent) more likely to take a graduation exam during their first year, with statistically significant evidence of increased exam passage as well. For the 2018 cohort in District B, assignment to the program is associated with a decline of 4.6 percentage points (5.8 percent) in exam-taking, while estimated effects on exam passage are negative and statistically insignificant. Effects for District A are smaller and not precisely estimated.

[15] For several outcomes, particularly chronic absenteeism in Year 1, passing at least one graduation exam, and dropout in Year 1, estimated effects are more favorable for District B Cohort 17 when using the full outcome-specific samples rather than restricting to students observed on all outcomes. Results using these less restrictive samples are reported in Appendix Tables A.15 - A.21 (analytic sample) and Appendix Tables A.22 - A.25 (full sample).

[16] Attendance data are missing for District B in the 2020-2021 academic year (Year 4 for the 2017 cohort and Year 3 for the 2018 cohort).

Table 4: ITT Effect of Program on Early Outcomes By District and Cohort

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
Credits in Year 1	-0.022	0.018	0.364*	-0.303
	(0.111)	–	(0.186)	(0.234)
	[4.888]	[5.432]	[4.380]	[4.714]
Credits Accumulated in Years 1 and 2	0.028	0.088	0.459	-0.313
	–	(0.229)	(0.382)	(0.234)
	[8.368]	[10.833]	[8.719]	[5.510]
Above Credit Threshold	-0.011	0.004	0.027	-0.045
	(0.019)	(0.020)	(0.039)	(0.038)
	[0.560]	[0.613]	[0.509]	[0.544]
Chronically Absent in Year 1	-0.023	-0.030	-0.058**	0.014
	(0.015)	(0.022)	(0.023)	(0.031)
	[0.361]	[0.463]	[0.359]	[0.263]
Chronically Absent in Years 1 or 2	-0.005	-0.029	-0.025	0.031
	–	(0.029)	(0.027)	(0.038)
	[0.506]	[0.616]	[0.455]	[0.434]
Took Graduation Exam	0.022	0.042	0.098***	-0.046
	(0.019)	–	(0.032)	(0.027)
	[0.633]	[0.392]	[0.747]	[0.789]
Passed at Least 1 Graduation Exam	0.016	0.033	0.066*	-0.041
	(0.018)	(0.028)	(0.033)	(0.030)
	[0.388]	[0.334]	[0.376]	[0.449]
Dropout in Year 1	-0.002	-0.002	-0.014**	0.009
	(0.003)	(0.002)	(0.006)	(0.007)
	[0.007]	[0.003]	[0.017]	[0.003]
Dropout in Year 2	0.003	-0.000	0.003	0.005
	(0.003)	(0.004)	(0.007)	(0.003)
	[0.003]	[0.003]	[0.008]	[0.000]
Dropout in Year 3	0.002	-0.001	0.002	0.001
	(0.005)	(0.009)	(0.011)	(0.008)
	[0.016]	[0.021]	[0.013]	[0.012]
Dropout in Year 4	-0.003	-0.002	-0.009	-0.005
	–	–	(0.010)	(0.007)
	[0.032]	[0.055]	[0.021]	[0.018]
On Track Version A	0.010	0.038	0.051	-0.057*
	(0.019)	(0.029)	(0.035)	(0.031)
	[0.386]	[0.324]	[0.386]	[0.447]
On Track Version B	0.014	0.041*	0.028	-0.036
	–	(0.023)	(0.032)	(0.036)
	[0.326]	[0.246]	[0.346]	[0.389]
On Track Version C	-0.000	0.025	0.038	-0.061
	–	(0.021)	(0.032)	(0.046)
	[0.574]	[0.495]	[0.588]	[0.641]
N	2312	916	670	726

Notes: All estimates represent intent-to-treat (ITT) effects, with treatment defined as ever being offered the program and the last complier dropped from each waitlist. All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, spoke English as a Second Language, in poverty, age 15 at enrollment, age 16 or over at enrollment, and whether they have more than two risk factors. Estimates are based on the analytic sample and, within each column, are restricted to students observed for all outcomes reported in that column to maintain a constant sample across outcomes. The resulting sample sizes can be found in the last row of the table. Results are qualitatively similar using the full outcome-specific samples, which can be found in the appendix (Tables A.15 through A.21). Standard errors from the bias-corrected sandwich (HC2) estimator are shown in parentheses. In a small share of cells, this standard error is left blank because it cannot be reliably computed, as discussed in footnote 12. *** p<0.01 ** p<0.05 * p<0.10. Control group means are shown in brackets.

The remaining early and intermediate outcomes we consider include persistence in school and indicators of being on track. We measure persistence using indicators for whether a student drops out by the end of each school year. Due in large part to compulsory schooling laws, dropout is rare in the first two years across all groups, and estimated treatment effects are generally small and statistically indistinguishable from zero. An exception is District B Cohort 2017, where assignment to the program reduced dropout in Year 1 by 1.4 percentage points, statistically significant at the 5 percent level, consistent with the broader pattern of positive early impacts in that cohort.

We construct composite on-track measures that aggregate multiple early warning indicators into a single summary of student trajectory. Consistent with prior work identifying early academic benchmarks as strong predictors of graduation (Kemple and Stephenson, 2013; Allensworth, 2005), we define three increasingly inclusive on-track measures using student-level administrative records. The first version classifies a student as off-track if they accumulated fewer than 2.5 credits in their first year or did not pass at least one required graduation exam in Year 1. The second version broadens the definition further by classifying students as off-track if any of four indicators are observed: (i) dropout in Year 1, (ii) fewer than 2.5 credits in Year 1, (iii) chronic absenteeism in Year 1, or (iv) not passing at least one graduation exam in Year 1. The third version omits the graduation exam component, classifying students as off-track based only on dropout, credit accumulation, and chronic absenteeism in Year 1. Across these composite measures, estimated effects for the 2017 cohort in District B are positive and consistent with the broader graduation results, though imprecisely estimated.

Taken together, differences in program impacts across districts and cohorts are already evident in early high school outcomes and follow a consistent pattern across multiple measures. For the cohort-district with the strongest graduation effects, improvements in credit accumulation, attendance, and exam participation appear by the end of ninth grade and persist over time, rather than emerging during the COVID-19 period or following pandemic-related policy changes. This timing makes it unlikely that differences in pandemic context alone explain the heterogeneity in graduation impacts. Instead, the close alignment between early outcomes and later graduation results points to differences in treatment timing, exposure, and engagement as key contributors.

VII Heterogeneity by Student Characteristics

The analysis above highlights meaningful differences in treatment effects across districts and cohorts, suggesting that differences in implementation and treatment exposure may have shaped program impacts. We now consider whether the participants themselves contributed to variation by exploring treatment effects for student subgroups. The broader literature on evaluating programs has increasingly emphasized the salience of treatment heterogeneity. For example, a recent essay by Smith (2022) goes so far as to argue "for the general empirical irrelevance of the common effect model." Consistent with this literature, our analysis considers the estimated impact of the CHSS program for subgroups. In Section 5.1, we highlighted sharp heterogeneity in treatment effects across cohorts and districts. To further understand the nature of heterogeneous impacts of the CHSS program, we also consider heterogeneity by demographic groups defined by gender, race, and ethnicity, the number of risk factors, and predicted graduation rate.[17][18]

Figure 3a reports the ITT estimates of program impact for several groups for the pooled sample.[19] The dot represents the regression-adjusted difference in graduation rates between the treatment and control groups. The horizontal lines through each dot represent the 95 percent confidence interval, calculated using bias-corrected sandwich (HC2) standard errors clustered by school, risk factor, gender, and cohort. At the top of the figure, we include the estimate for all students, which is the same as the estimate reported in column 1 of Table 2. These results for the pooled sample show little evidence of an effect of the program for most demographic groups. An exception is students with disabilities, for whom the estimated effect is positive and statistically significant. This effect is notable because it is present for both the 2017 and 2018 cohorts in District B (special education status was not available for students

[17] As pre-registered, we compute these predicted graduation values using the repeated split-sample procedure described in Abadie et al. (2018). Specifically, for each repetition, we use half of the control group to estimate a prediction model using baseline characteristics and school fixed effects, and apply the resulting coefficients to the full sample. We average predictions across repetitions to obtain a prediction index that avoids endogenous stratification.

[18] Pre-registered subgroup analyses include gender, race and ethnicity, age, predicted likelihood of graduation, and baseline risk factors, which we specified in advance as confirmatory tests of treatment effect heterogeneity. We additionally test for effects by students with disabilities in the two cohorts where these data are available, though this analysis was not pre-registered. Analyses disaggregated by district and cohort, as well as those examining other student characteristics such as disability status, were not pre-specified and are therefore exploratory. We also present exploratory evidence on longer-run post-secondary outcomes in Appendix B.

[19] Corresponding subgroup analyses for the early academic outcomes discussed in Section 6.3 are presented as ladder plots in Appendix Figures A.1 - A.4.

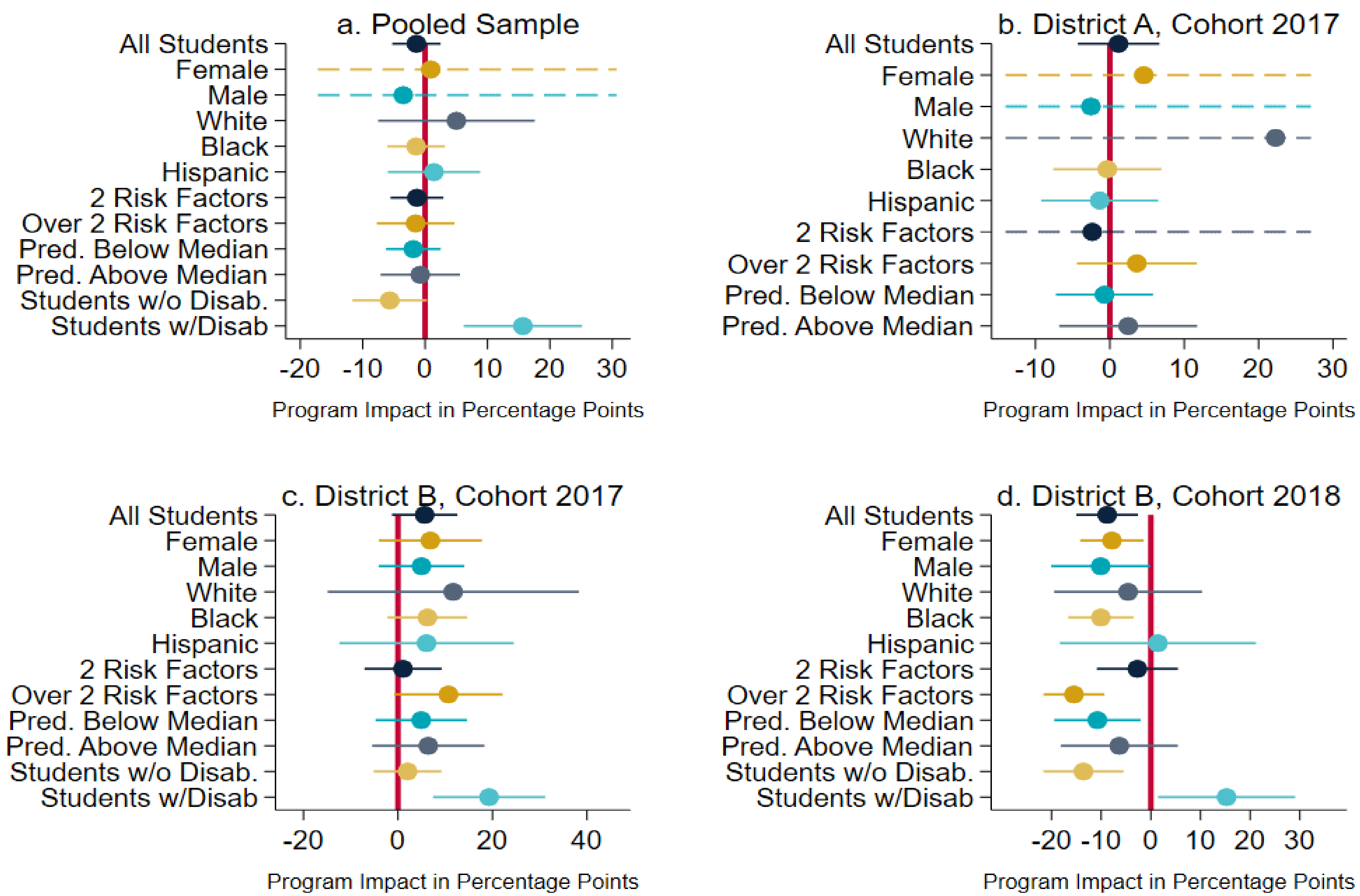
in District A), suggesting particularly concentrated benefits of program participation for this group of students. There are 292 students with disabilities, or 21 percent of the District B analytic sample. Among control group students with disabilities, 59 percent graduated within four years, compared with roughly 80 percent for control group students without disabilities. For White students, the point estimate suggests that the program increased graduation rates, but this estimate is very imprecise.

We observe differences in treatment effects across demographic groups within the district and cohort subsamples.[20] Figures 3.b-3.d report the results separately by district and cohort. An exception is White students, for whom the estimated effect is positive and statistically significant, though this result should be interpreted with caution given that White students comprise only about 10 percent of the District A sample. For the District B 2017 cohort in 3.c, the treatment effect is positive for all groups. While most of these estimates are imprecise, the treatment effect is positive for the most at-risk students, those with the lowest predicted likelihood of graduating high school based on their baseline characteristics. For this group, those assigned to treatment are 4.9 percentage points more likely to graduate within four years than those assigned to the control group, though this estimate is not statistically significant at conventional levels. For the 2018 District B Cohort in 3.d, the point estimate is negative for nearly all subgroups, and for more than half of them, this effect is statistically significant. The negative effect is particularly evident for Black students, students with more than two risk factors, and students without disabilities.

While exploratory, these subgroup differences may partially reflect variation in treatment take-up across student groups. Appendix Table A.3 indicates that assignment to treatment generated particularly large enrollment responses for Hispanic students, female students, and students with exactly two risk factors. These patterns broadly align with the treatment effects observed for some of these groups in Figure 3, though differences in take-up alone are unlikely to fully explain the heterogeneity in impacts.

[20]We also tested for heterogeneity by above-median age at enrollment (pre-registered) and English Language Learner (ELL) status (not pre-registered). Results are not reported because there is little variation in treatment effects across these groups.

Figure 3: Intent-to-Treat Estimates of the Effect of the Program on High School Graduation by District, Cohort, and Student Characteristics



Notes: Each point is the regression-adjusted ITT difference in four-year graduation between the treatment and control groups, using the ever-offer treatment definition and dropping the last complier from each waitlist to account for the endogenous stopping rule (de Chaisemartin and Behaghel, 2020). All estimates include school and cohort fixed effects. Standard errors are computed using the reg_sandwich procedure with HC2 bias-corrected inference, clustered at the school/risk-factor-category/gender/cohort level. Bars represent 95% confidence intervals. In a small share of subgroups, the confidence interval cannot be reliably computed and the bars are represented as dashed horizontal lines spanning the figures. Footnote 12 discusses this computation decision. "Pred." denotes predicted graduation rate, constructed using baseline covariates following the repeated split-sample procedure of Abadie et al. (2018); the "Pred. Below Median" bars represent program impacts for students whose predicted graduation probability falls below the sample median. "Students w/o Disab." denotes students without disabilities and "Students w/Disab." denotes students with disabilities. District A did not provide information on student disability status.

VIII Conclusion

Interventions that integrate comprehensive supports and services in order to improve outcomes for at-risk high school students have attracted considerable attention in recent years. In this study, we examine the impact of such an intervention through a large-scale randomized controlled trial. Overall, we find that random assignment to the CHSS program had small and statistically insignificant effects on the primary outcome of interest, high school graduation. However, we find noticeable differences in the treatment effect across cohorts and school districts. Using program data on enrollment and the nature of services offered, we find extraordinary heterogeneity in program implementation across district-year cohorts that appears to be related to the effect of the program–we find evidence of a positive treatment effect for the cohort and district where students were enrolled early on in high school and where treatment intensity was high. In contrast, we find little evidence of an effect of the program in the district-cohort where students were enrolled later in the academic year and when treatment intensity was lighter on a number of dimensions. In the district-cohort where enrollment was particularly late and treatment intensity was lightest, we found evidence of negative effects on high school completion.

These results demonstrate that average treatment effects of comprehensive case management programs, offering wraparound supports to high school students, are fragile. While comprehensive models have the potential to improve graduation outcomes, their success depends critically on local implementation capacity and fidelity. Even well-designed programs may fail to generate consistent effects in the absence of timely recruitment, staff continuity, and service intensity. In this sense, the benefits of wraparound interventions are contingent on fidelity to a flagship model, with realized effects hinging on whether schools can sustain high-quality, early, and coordinated service delivery.

The results also highlight the limited generalizability of much of the existing evidence on programs to reduce high school dropout. Single-site or single-cohort randomized controlled trials may offer compelling causal estimates in context, but these can obscure meaningful variation across settings and years. In the case of CHSS, the same program produced large positive effects in one cohort and negative effects in another *within the same district*. Such divergent findings underscore the need for multiple replications across time and place before general claims of effectiveness are made. This pattern is not unique to CHSS. Prior randomized evaluations of Check & Connect, Early College, and Career Academies likewise exhibit wide variation in estimated impacts across sites and cohorts, underscoring the broader challenges of generalizing findings from isolated settings.

We believe the study carries important policy implications. Effective school support policy requires more than identifying promising models; it requires rigorous, localized evaluation and careful attention to scalability. Policymakers should recognize that expanding comprehensive support programs entails significant risks of implementation slippage, particularly in resource-constrained districts. Investments in wraparound services should therefore be coupled with mechanisms to monitor fidelity, adapt to local conditions, and continuously improve delivery if the promise of these interventions is to be realized.

References

Alberto Abadie, Matthew M. Chingos, and Martin R. West. Endogenous stratification in randomized experiments. *Review of Economics and Statistics*, 2018.

Alberto Abadie, Susan Athey, Guido W. Imbens, and Jeffrey M. Wooldridge. When should you adjust standard errors for clustering? *Quarterly Journal of Economics*, 138:1–35, 2023.

Melissa Adelman, Francisco Haimovich, Mauricio Romero, and Emmanuel Vazquez. Preventing school dropout at scale: Experimental evidence from guatemala. *Journal of Labor Economics*, 2025.

Hunt Allcott. Site selection bias in program evaluation. *The Quarterly Journal of Economics*, 130(3):1117–1165, 2015. doi: 10.1093/qje/qjv015.

Elaine M. Allensworth. Graduation and dropout trends in chicago: A look at cohorts of students from 1991 through 2004. Technical report, Consortium on Chicago School Research at the University of Chicago, 2005.

Elaine M. Allensworth and John Q. Easton. The on-track indicator as a predictor of high school graduation. Technical report, Consortium on Chicago School Research, 2005.

Andrea Berger, Lori Turk-Bicakci, Michael Garet, Joel Knudson, and Gabe Hoshen. Early college, early success: Early college high school initiative impact study. Technical report, American Institutes for Research, 2013.

David Card. The causal effect of education on earnings. *Handbook of Labor Economics*, 1999.

Julie Berry Cullen, Steven D. Levitt, Erin Robertson, and Sally Sadoff. What can be done to improve struggling high schools? *Journal of Economic Perspectives*, 2013.

Jonathan M. V. Davis, Jonathan Guryan, Kelly Hallberg, and Jens Ludwig. The economics of scale-up. NBER Working Paper 23925, National Bureau of Economic Research, 2017.

Clément de Chaisemartin and Luc Behaghel. Estimating the effect of treatments allocated by randomized waiting lists. *Econometrica*, 2020.

Thomas S. Dee. The promises and the reality of school reform: Three forces that matter. Policy brief, Hoover Institution, 2024.

Mark Dynarski, Linda Clarke, Brian Cobb, Jeremy Finn, Russell Rumberger, and Jay Smink. Dropout prevention. ies practice guide. Technical Report NCEE 2008–4025, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2008.

Julie A. Edmunds, Lora Bernstein, Fatih Unlu, Elizabeth Glennie, Abby Smith, and Nina Arshavsky. Smoothing the transition to postsecondary education: The impact of the early college model. *Journal of Research on Educational Effectiveness*, 10(2):297–325, 2017. doi: 10.1080/19345747.2016.1191574.

Jonathan Guryan, Sandra Christenson, Ashley Cureton, Ijun Lai, Jens Ludwig, Catherine Schwarz, Emma Shirey, and Mary Clair Turner. The effect of mentoring on school attendance and academic outcomes: A randomized evaluation of the check connect program. *Journal of Policy Analysis and Management*, 2020.

Sara Heller, Harold A. Pollack, Roseanna Ander, and Jens Ludwig. Preventing youth violence and dropout: A randomized field experiment. *American Economic Review*, 103(3):311–315, 2013.

Sara B. Heller, Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A. Pollack. Thinking, fast and slow? some field experiments to reduce crime and dropout in chicago. *Quarterly Journal of Economics*, 132(1):1–54, 2017. doi: 10.1093/qje/qjw033.

Jessica B. Heppen, Kristina Zeiser, Deborah J. Holtzman, Madeline O'cummings, Sandra Christenson, and Alexandra Pohl. Efficacy of the check & connect mentoring program for at-risk general education high school students. *Journal of Research on Educational Effectiveness*, 11(1):56–82, 2018.

James J. Kemple. Career academies: Impacts on students' initial transitions to post-secondary education and employment. Technical report, Manpower Demonstration Research Corporation, 2001.

Micha D. Segeritz Kemple, James J. and Nickisha Stephenson. Building on-track indicators for high school graduation and college readiness: Evidence from new york city. *Journal of Education for Students Placed at Risk*, 18:7–28, 2013.

Matthew A. Kraft. Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4):241–253, 2020. doi: 10.3102/0013189X20912798.

John A. List. *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale*. Crown Currency, New York, 2022.

Brandy R. Maynard, Elizabeth K. Kjellstrand, and Aaron M. Thompson. Effects of check and connect on attendance, behavior and academics: A randomized effectiveness trial. *Research on Social Work Practice*, 24:296–309, 2014.

National Center for Education Statistics. Digest of education statistics, table 219.10: High school graduates, by sex and control of school; public high school averaged freshman graduation rate (afgr); public high school 4-year adjusted cohort graduation rate (acgr); and total graduates as a ratio of 17-year-old population: Selected school years, 1869-70 through 2022-23. Online, 2025. URL `https://nces.ed.gov/programs/digest/d24/tables/dt24_219.10.asp?current=yes`. Accessed July 14, 2025.

National Student Clearinghouse. About the national student clearinghouse. `https://www.studentclearinghouse.org/about/`, 2025. Accessed September 28, 2025.

Russell W. Rumberger, Howard Addis, Elaine Allensworth, Robert Balfanz, Debra Duardo, and Mark Dynarski. Dropout prevention: A practice guide. Technical Report NCEE 2017-4028, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, 2017.

Mary F. Sinclair, Sandra L. Christenson, and Martha L. Thurlow. Dropout prevention for youth with disabilities: Efficacy of a sustained school engagement procedure. *Exceptional Children*, 65(1):7–21, 1998.

Mary F. Sinclair, Sandra L. Christenson, and Martha L. Thurlow. Promoting school completion of urban secondary youth with emotional or behavioral disabilities. *Exceptional Children*, 71(4):465–482, 2005.

Jeffrey Smith. Treatment effect heterogeneity. *Evaluation Review*, 2022.

Michael J. Weiss, Howard S. Bloom, and Thomas Brock. A conceptual framework for studying the sources of variation in program effects. *Journal of Policy Analysis and Management*, 2014.

A Appendix A

Table A.1: Regression Adjusted Differences in Sample Composition and Outcome Availability by Treatment Status

	Full Sample	Control	Treatment	Difference
	Means	Means	Means	Treatment - Control
Missing Grad Outcome	0.0586	0.0577	0.0592	0.00094
				(0.0097)
Transferred Outside of State	0.0179	0.0241	0.0142	-0.0091
				(0.0058)
Unknown Reason for Missingness	0.0407	0.0336	0.0450	0.010
				(0.0079)
Transferred to HSE Program	0.0399	0.0357	0.0425	0.011
				(0.0079)
Match to NSC	0.989	0.993	0.987	-0.0045
				(0.0037)
In Analytic Sample	0.941	0.942	0.941	-0.00094
				(0.0097)
Match to NSC and in Analytic Sample	0.937	0.939	0.936	-0.0035
				(0.0100)
Total Obs.	2,507	953	1,554	2,507

Notes: The difference column reports regression-adjusted differences between treatment and control groups, using ever-offer as the treatment definition. Control variables include indicators for whether a student is female and whether the student has more than two risk factors. Missingness comparisons are conducted using the full randomized sample. Missingness is defined as having transferred to a school outside of the state or having unknown or otherwise ambiguous enrollment statuses. Exact matches on first name, last name, and date of birth were required to successfully match students to National Student Clearinghouse (NSC) records. Standard errors reported in parentheses are heteroskedasticity-robust. *** $p<0.01$ ** $p<0.05$ * $p<0.10$.

Table A.2: Regression Adjusted Differences in Sample Composition and Outcome Availability by Treatment Status, District, and Cohort

	Control	Treatment	Total (District A)	Control	Treatment	Total (District B C17)	Control	Treatment	Total (District B C18)
	Means	Means	Treatment - Control	Means	Means	Treatment - Control	Means	Means	Treatment - Control
Missing Grad Outcome	0.0436	0.0585	0.015	0.0669	0.0604	-0.0067	0.0648	0.0590	-0.0079
			(0.014)			(0.019)			(0.017)
Transferred Outside of State	0.00291	0.00154	-0.0013	0	0	0	0.0620	0.0495	-0.012
			(0.0033)			–			(0.017)
Unknown Reason for Missingness	0.0407	0.0569	0.016	0.0669	0.0604	-0.0067	0.00282	0.00943	0.0042
			(0.014)			(0.019)			(0.0049)
Transferred to HSE Program	0.0116	0.0123	0.00037	0.0551	0.0521	-0.0025	0.0451	0.0778	0.033*
			(0.0072)			(0.017)			(0.017)
Match to NSC	0.980	0.971	-0.012	1	1	0	1	0.998	0
			(0.0097)			–			–
In Analytic Sample	0.956	0.942	-0.015	0.933	0.940	0.0067	0.935	0.941	0.0079
			(0.014)			(0.019)			(0.017)
Match to NSC and in Analytic Sample	0.948	0.929	-0.021	0.933	0.940	0.0067	0.935	0.941	0.0079
			(0.015)			(0.019)			(0.017)
Total Obs.	344	650	994	254	480	734	355	424	779

Notes: The difference columns report regression-adjusted differences between treatment and control groups, using ever-offer as the treatment definition. Control variables include indicators for whether a student is female and whether the student has more than two risk factors. Missingness comparisons are conducted using the full randomized sample. Missingness is defined as having transferred to a school outside of the state or having unknown or otherwise ambiguous enrollment statuses. Standard errors reported in parentheses are heteroskedasticity-robust. *** p<0.01 ** p<0.05 * p<0.10.

Table A.3: Program Take-Up by District, Cohort, and Student Characteristics

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(1)	(2)	(3)	(4)
(a) Full Sample				
Ever Offer	0.364	0.400	0.349	0.337
	(0.026***)	(0.037***)	(0.044***)	(0.058***)
F-Statistic	200.57	117.34	63.54	33.78
N	2,360	941	688	731
(b) First Stage by Subgroup				
All Students	0.364***	0.400***	0.349***	0.337***
	(0.026)	(0.037)	(0.044)	(0.058)
Female	0.415	0.463	0.395***	0.400***
	–	–	(0.066)	(0.067)
Male	0.316	0.338	0.305***	0.281**
	–	–	(0.058)	(0.092)
White	0.246***	0.299**	0.225***	0.283***
	(0.045)	(0.109)	(0.070)	(0.078)
Black	0.397***	0.403***	0.400***	0.374***
	(0.027)	(0.036)	(0.051)	(0.063)
Hispanic	0.439***	0.506***	0.355***	0.319***
	(0.044)	(0.065)	(0.067)	(0.087)
2 Risk Factors	0.434***	0.477	0.418***	0.412***
	(0.030)	–	(0.053)	(0.069)
Over 2 Risk Factors	0.301***	0.335***	0.283***	0.276**
	(0.036)	(0.043)	(0.064)	(0.088)
Students w/o Disabilities	0.367***	–	0.367***	0.363***
	(0.040)	–	(0.047)	(0.064)
Students w/ Disabilities	0.274***	–	0.306***	0.251***
	(0.046)	–	(0.060)	(0.065)

Notes: Estimates are intention-to-treat (ITT) first-stage effects estimated via OLS, with treatment defined using the ever-offer definition and the last complier dropped from each waitlist. Panel (a) reports the first stage for the full analytic sample. Panel (b) reports the first stage estimated separately for each subgroup listed. All regressions include school and cohort fixed effects. Disability status was not available for District A. Standard errors in parentheses are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level. In a small share of cells, the clustered HC2 standard error is left blank because it cannot be reliably computed, as discussed in footnote 12. *** p<0.01 ** p<0.05 * p<0.10.

Table A.4: Sensitivity of Graduation Effects to Alternative Standard Error Clustering

	(1)	(2)	(3)	(4)
	Pooled	District A C17	District B C17	District B C18
Graduated within 4 Years				
Ever Offer	-0.014	0.012	0.057	-0.088
	(0.0177)	(0.0274)	(0.0333*)	(0.0319***)
	[0.0202]	[0.0291]	[0.0155**]	[0.0209**]
	((0.0193))	((0.0261))	((0.0354))	((0.0316**))
	[[0.0198]]	[[0.0280]]	[[0.0353]]	[[0.0317**]]
	|0.0179|	|0.0283|	|0.0338*|	|0.0328***|
	<0.0129>	<0.0268>	<0.0157***>	<0.0191***>
	/0.0196/	/0.0253/	/0.0342*/	/0.0309***/
Control Mean	0.7294	0.7143	0.7257	0.7470
N	2,360	941	688	731

Notes: Coefficients from Table 2 are shown in the first row, with treatment defined using the ever-offer definition and the last complier dropped from each waitlist. All regressions include school and cohort fixed effects. All other values are standard errors computed under alternative inference procedures. ( ): unclustered (robust) standard errors [ ]: school-level clustering (( )): school/risk factor/gender clustering [[ ]]: bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level (pre-registered specification) | |: bootstrapped unclustered < >: bootstrapped school-level clustering / /: bootstrapped + clustered at the school/risk factor/gender level. *** p<0.01 ** p<0.05 * p<0.10.

Table A.5: Effect of Program on 4 and 5-Year Graduation, Pooled and by District

	Pooled		District A		District B			
	ITT	TOT	ITT	TOT	ITTC17	TOTC17	ITTC18	TOTC18
	(a) 4-Year Graduation							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ever Offer	-0.014		0.012		0.057		-0.088***	
	(0.020)		(0.028)		(0.035)		(0.032)	
Enrolled into Treatment		-0.038		0.029		0.162*		-0.261***
		(0.049)		(0.068)		(0.095)		(0.097)
Control Mean	.7294		.7143		.7257		.747	
Control Complier Mean		.8979		.8254		.7209		1.103
N	2360	2360	941	941	688	688	731	731
	(b) 5-Year Graduation							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ever Offer	-0.006		0.018		0.060*		-0.074*	
	(0.020)		–		(0.035)		(0.040)	
Enrolled into Treatment		-0.016		0.046		0.171*		-0.220**
		(0.047)		(0.066)		(0.094)		(0.093)
Control Mean	.7572		.7508		.7342		.7801	
Control Complier Mean		.8905		.8293		.7118		1.083
N	2360	2360	941	941	688	688	731	731

Notes: Estimates represent intention-to-treat (ITT) and treatment-on-the-treated (TOT) effects, with treatment defined as ever being offered the program and the final treated student in each waitlist excluded to account for the endogenous stopping rule in treatment offers (de Chaisemartin and Behaghel, 2020). TOT estimates are computed using two-stage least squares, instrumenting program enrollment with treatment assignment. Pooled specifications include school and cohort fixed effects; district-specific specifications include school fixed effects only. Control variables include indicators for whether a student is female, Black, White, Hispanic, English as a Second Language, in poverty, age 15 at enrollment, age 16 or older at enrollment, and whether the student has more than two risk factors. Estimates are based on the analytic sample of 2,360 students with observable graduation outcomes, after excluding the last complier in each waitlist. ITT standard errors in parentheses are computed using the reg_sandwich procedure with bias-corrected sandwich (HC2) inference and clustering at the cohort/school/risk-factor-category/gender level. In a small share of cells, the clustered HC2 standard error is left blank because it cannot be reliably computed, as discussed in footnote 12. *** p<0.01 ** p<0.05 * p<0.10. The control complier mean reflects the average outcome for students in the control group who would have enrolled if offered the program, calculated as the mean outcome for treatment compliers minus the estimated TOT effect. The control complier mean for District B Cohort 2018 exceeds 1 due to the combination of a large negative TOT estimate and modest first-stage take-up rate, and should not be interpreted as a graduation probability.

Table A.6: Baseline Characteristics of Program Compliers

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(1)	(2)	(3)	(4)
	(a) Full Sample			
White	**0.099**	0.040	0.127	0.161
Black	**0.599**	0.599	0.627	0.568
Hispanic	**0.234**	0.348	0.157	0.135
English Second Language	**0.229**	0.308	0.133	0.206
In Poverty	**0.938**	0.964	0.886	0.955
Enrollment Age (Years)	**14.03**	14.10	14.00	13.96
Predicted Graduation Rate	**0.777**	0.752	0.783	0.809
Female	**0.551**	0.555	0.542	0.555
Over Two Risk Factors	**0.440**	0.478	0.373	0.452
Students w/Disabilities	**0.182**	–	0.211	0.145
Total Obs.	**568**	247	166	155
	(b) Analytic Sample			
White	**0.096**	0.041	0.116	0.162
Black	**0.604**	0.602	0.634	0.574
Hispanic	**0.231**	0.344	0.159	0.128
English Second Language	**0.221**	0.299	0.122	0.203
In Poverty	**0.937**	0.963	0.884	0.953
Enrollment Age (Years)	**14.03**	14.09	13.99	13.97
Predicted Graduation Rate	**0.777**	0.753	0.783	0.809
Female	**0.553**	0.556	0.549	0.554
Over Two Risk Factors	**0.438**	0.469	0.378	0.453
Students w/Disabilities	**0.184**	–	0.213	0.146
Total Obs.	**553**	241	164	148

Notes: Each cell reports the sample mean of the row variable within the indicated column, restricted to program compliers (students who ever enrolled), dropping the last complier from each waitlist. Panel (a) covers the full sample; panel (b) restricts to the analytic sample of students with observable graduation outcomes. Age at enrollment is reported in years. All other rows are proportions except the predicted graduation rate. District A did not provide information on student disability status.

Table A.7: Robustness to Treatment Definition

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(a) 4-Year Graduation			
	(1)	(2)	(3)	(4)
Initial Offer	**-0.012**	0.011	0.051	-0.085
	–	–	(0.034)	(0.044*)
	[0.018]	[0.028]	[0.033]	[0.033**]
Control Mean	**0.7341**	0.7130	0.7308	0.7654
N	**2,360**	941	688	731
	(b) 5-Year Graduation			
	(1)	(2)	(3)	(4)
Initial Offer	**-0.004**	0.021	0.054	-0.074
	–	–	(0.034)	(0.045)
	[0.018]	[0.027]	[0.033]	[0.032**]
Control Mean	**0.7603**	0.7469	0.7393	0.7984
N	**2,360**	941	688	731

Notes: Estimates are ITT effects, with treatment defined using the initial-offer definition and the last complier dropped from each waitlist to account for the endogenous stopping rule in randomized waitlist designs (de Chaisemartin and Behaghel, 2020). All estimates include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, spoke English as a Second Language, in poverty, age 15 at enrollment, age 16 or over at enrollment, and whether they have more than two risk factors. After dropping the last complier in each waitlist, the analytic sample includes 2,360 students with observable graduation outcomes. Standard errors in parentheses ( ) are robust unclustered standard errors; those in brackets [ ] are clustered by school/risk factor/gender/cohort. *** p<0.01 ** p<0.05 * p<0.10. Bolded analyses were pre-registered.

Table A.8: ITT Effect on High School Graduation Using the Doubly Reweighted Ever-Offer Estimator

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(a) 4-Year Graduation			
	(1)	(2)	(3)	(4)
Doubly-Reweighted Ever Offer	**-0.012**	0.012	0.055	-0.087
	(0.020)	(0.030)	(0.012**)	(0.018***)
Control Mean	**0.7294**	0.7143	0.7257	0.7470
N	**2,381**	952	693	736
	(b) 5-Year Graduation			
	(1)	(2)	(3)	(4)
Doubly-Reweighted Ever Offer	**-0.004**	0.019	0.059	-0.074
	(0.019)	(0.021)	(0.018**)	(0.034*)
Control Mean	**0.7572**	0.7508	0.7342	0.7801
N	**2,381**	952	693	736

Notes: Estimates are intention-to-treat (ITT) effects computed using the Doubly Reweighted Ever-Offer (DREO) estimator for randomized waitlists (de Chaisemartin and Behaghel, 2020). Control variables include binary indicators for whether a student is female, Black, White, Hispanic, spoke English as a Second Language, in poverty, age 15 at enrollment, age 16 or over at enrollment, and whether they have more than two risk factors. The analytic sample includes 2,381 students with observable graduation outcomes. Standard errors in parentheses are computed using the randomization-based variance estimator proposed by de Chaisemartin and Behaghel (2020), with degrees of freedom equal to the number of waitlists minus one. *** p<0.01 ** p<0.05 * p<0.10. Bolded analyses were pre-registered.

Table A.9: Baseline Balance Table by District and Cohort, Full Sample

	Control	Treatment	Total (District A)	Control	Treatment	Total (District B C17)	Control	Treatment	Total (District B C18)
	Means	Means	Treatment - Control	Means	Means	Treatment - Control	Means	Means	Treatment - Control
White	0.0962	0.0708	-0.026	0.228	0.204	-0.024	0.208	0.227	0.019
			(0.019)			(0.032)			(0.030)
Black	0.528	0.575	0.047	0.539	0.519	-0.020	0.493	0.501	0.0080
			(0.033)			(0.039)			(0.036)
Hispanic	0.321	0.294	-0.027	0.118	0.163	0.044*	0.141	0.135	-0.0061
			(0.031)			(0.026)			(0.025)
English Second Language	0.327	0.309	-0.018	0.209	0.204	-0.0052	0.276	0.246	-0.030
			(0.031)			(0.031)			(0.032)
In Poverty	0.950	0.948	-0.0041	0.898	0.908	0.011	0.935	0.946	0.011
			(0.014)			(0.023)			(0.017)
Enrollment Age (Years)	14.13	14.18	0.035	14.09	14.06	-0.025	14.11	14.01	-0.094**
			(0.040)			(0.049)			(0.042)
Students w/Disabilities	–	–	–	0.201	0.223	0.023	0.219	0.214	-0.0020
			(0.040)			(0.031)			(0.032)
Predicted Graduation Rate	0.713	0.707	0.00011	0.726	0.725	-0.0033	0.746	0.758	0.0099
			(0.012)			(0.0085)			(0.0077)
Female	0.478	0.477	0.00039	0.476	0.483	0.0060	0.462	0.468	0.0050
			(0.033)			(0.039)			(0.036)
Over Two Risk Factors	0.539	0.560	0.021	0.488	0.479	-0.0083	0.546	0.539	-0.0065
			(0.033)			(0.039)			(0.035)
Total Obs.	344	650	994	254	480	734	355	424	779

Notes: The difference columns report regression-adjusted differences between treatment and control groups, with ever-offer used as the treatment definition and the final treated student in each waitlist excluded to account for the endogenous stopping rule in treatment offers (de Chaisemartin and Behaghel, 2020). The predicted graduation rate is a predicted value constructed from the other baseline covariates in the table following the repeated split-sample (RSS) procedure of Abadie et al. (2018). Disability status was not available for District A. Standard errors reported in parentheses are heteroskedasticity-robust. *** $p<0.01$ ** $p<0.05$ * $p<0.10$.

Table A.10: Baseline Balance Table by District and Cohort, Analytic Sample

	Control Means	Treatment Means	Total (District A) Treatment - Control	Control Means	Treatment Means	Total (District B C17) Treatment - Control	Control Means	Treatment Means	Total (District B C18) Treatment - Control
White	0.101	0.0670	-0.034*	0.224	0.193	-0.030	0.214	0.231	0.017
			(0.019)			(0.033)			(0.031)
Black	0.530	0.587	0.056	0.553	0.532	-0.021	0.512	0.506	-0.0072
			(0.034)			(0.040)			(0.037)
Hispanic	0.323	0.284	-0.039	0.114	0.166	0.052*	0.130	0.133	0.0036
			(0.032)			(0.027)			(0.025)
English Second Language	0.311	0.297	-0.014	0.215	0.182	-0.033	0.247	0.226	-0.020
			(0.032)			(0.032)			(0.032)
In Poverty	0.948	0.949	-0.000075	0.903	0.909	0.0059	0.931	0.942	0.011
			(0.015)			(0.023)			(0.018)
Enrollment Age (Years)	14.13	14.16	0.020	14.08	14.05	-0.035	14.10	14.01	-0.094**
			(0.040)			(0.051)			(0.043)
Students w/Disabilities	–	–	–	0.211	0.228	0.018	0.221	0.210	-0.0088
			(0.040)			(0.033)			(0.032)
Predicted Graduation Rate	0.715	0.711	0.0024	0.727	0.726	-0.00089	0.747	0.759	0.011
			(0.012)			(0.0088)			(0.0080)
Female	0.485	0.482	-0.00090	0.477	0.483	0.0069	0.458	0.481	0.024
			(0.034)			(0.040)			(0.037)
Over Two Risk Factors	0.537	0.557	0.020	0.485	0.488	0.0033	0.539	0.544	0.0085
			(0.034)			(0.040)			(0.037)
Total Obs.	329	612	941	237	451	688	332	399	731

Notes: The difference column reports Treatment – Control from a regression of the covariate on the treatment indicator, with ever-offer used as the treatment definition and the final treated student in each waitlist excluded to account for the endogenous stopping rule in treatment offers (de Chaisemartin and Behaghel, 2020). The predicted graduation rate is constructed using the other covariates reported in the table. Disability status was not available for District A. The sample is restricted to the analytic sample of students with observable graduation outcomes. Standard errors reported in parentheses are heteroskedasticity-robust. *** $p<0.01$ ** $p<0.05$ * $p<0.10$.

Table A.11: Effects of Waitlist Position on Take-Up and Completion Outcomes

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(1)	(2)	(3)	(4)
(a) First Stage: Effect of Top of Waitlist on Program Take-Up				
Top of Waitlist	0.142***	0.195***	0.185***	0.019
	(0.032)	(0.042)	(0.044)	(0.063)
F-statistic	20.35	21.76	17.39	0.09
Take-Up Mean	0.3629	0.3800	0.3417	0.3608
N	1,554	650	480	424
(b) ITT Effect of Top of Waitlist on HSE Transfer				
Top of Waitlist	-0.031**	-0.007	-0.038	-0.046
	(0.014)	(0.013)	(0.028)	(0.035)
Control Mean	0.0425	0.0123	0.0521	0.0778
N	1,554	650	480	424
(c) ITT Effect of Top of Waitlist on 4-Year Graduation				
Top of Waitlist	0.036	-0.020	0.087	0.073
	–	(0.044)	(0.057)	(0.048)
Control Mean	0.7223	0.7173	0.7761	0.6692
N	1,462	612	451	399

Notes: Estimates are ITT effects restricted to treated students in all panels. Top of Waitlist is defined as having a waitlist number at or below the position of the last student to enroll in the program within the first semester (by January of freshman year). All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, in poverty, age at enrollment, and whether they have more than two risk factors. Standard errors in parentheses are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level. In a small share of cells, the clustered HC2 standard error is left blank because it cannot be reliably computed, as discussed in footnote 12. *** p<0.01 ** p<0.05 * p<0.10.Take-Up Mean reflects the share of treated students who enrolled in the program.

Table A.12: Early Academic Outcomes for HSE and Non-HSE Students

	All	Non-HSE	HSE	Adj. Difference
	Means	Means	Means	HSE Minus Non-HSE
	(a) Full Sample			
Credits in Year 1	4.877	5.025	1.319	-3.29***
				(0.20)
Above Credit Threshold	0.551	0.572	0.0612	-0.45***
				(0.029)
Chronically Absent in Year 1	0.361	0.349	0.643	0.25***
				(0.049)
Any Chronic Absenteeism in Years 1 and 2	0.514	0.499	0.869	0.31***
				(0.036)
Took Graduation Exam	0.625	0.634	0.400	-0.18***
				(0.050)
Passed at Least 1 Graduation Exam	0.381	0.393	0.0900	-0.24***
				(0.031)
Total Obs.	2,507	2,407	100	2,507
	(b) Analytic Sample			
Credits in Year 1	4.895	5.053	1.319	-3.30***
				(0.20)
Above Credit Threshold	0.554	0.576	0.0612	-0.45***
				(0.029)
Chronically Absent in Year 1	0.357	0.344	0.643	0.25***
				(0.049)
Any Chronic Absenteeism in Years 1 and 2	0.511	0.495	0.869	0.31***
				(0.036)
Took Graduation Exam	0.637	0.647	0.400	-0.19***
				(0.050)
Passed at Least 1 Graduation Exam	0.390	0.403	0.0900	-0.24***
				(0.031)
Total Obs.	2,360	2,260	100	2,360

Notes: The difference column reports the regression-adjusted difference in means between HSE and non-HSE students, adjusting for whether the student is female and whether the student had over two risk factors, with ever-offer used as the treatment definition and the final treated student in each waitlist excluded to account for the stopping rule inherent in randomized waitlist designs (de Chaisemartin and Behaghel, 2020). The full sample includes all students who were randomized for the study; the analytic sample restricts to students with observable graduation outcomes. Standard errors in parentheses are heteroskedasticity-robust. *** $p<0.01$ ** $p<0.05$ * $p<0.10$.

Table A.13: Effects of Waitlist Position on Take-Up and Completion Outcomes Using Alternate Definition of Top of Waitlist

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(1)	(2)	(3)	(4)
(a) First Stage: Effect of Top of Waitlist on Program Take-Up				
Top of Waitlist	0.139***	0.207***	0.112	0.021
	(0.037)	(0.042)	(0.069)	(0.070)
F-statistic	14.42	24.05	2.60	0.09
Take-Up Mean	0.3629	0.3800	0.3417	0.3608
N	1,554	650	480	424
(b) ITT Effect of Top of Waitlist on HSE Transfer				
Top of Waitlist	-0.020	0.008	-0.075*	0.006
	(0.016)	–	(0.036)	(0.040)
Control Mean	0.0425	0.0123	0.0521	0.0778
N	1,554	650	480	424
(c) ITT Effect of Top of Waitlist on 4-Year Graduation				
Top of Waitlist	0.053	0.011	0.126**	0.011
	–	(0.039)	(0.056)	(0.087)
Control Mean	0.7223	0.7173	0.7761	0.6692
N	1,462	612	451	399

Notes: Estimates are restricted to treated students in all panels. Top of Waitlist is defined as having a waitlist number at or below the median waitlist rank within each school-by-cohort waitlist. All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, in poverty, age at enrollment, and whether they have more than two risk factors. Standard errors in parentheses are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level. In a small share of cells, the clustered HC2 standard error is left blank because it cannot be reliably computed, as discussed in footnote 12. *** p<0.01 ** p<0.05 * p<0.10. Take-Up Mean reflects the share of treated students who enrolled in the program.

Table A.14: Program Activity Participation Rates Among Enrollees, by District and Cohort

	District A C17	District B C17	District B C18	Dist. A C17 minus Dist. B C17	Dist. B C18 minus Dist. B C17
	Means			Means (Standard Errors)	
Tutoring	0.667	0.842	0.600	-0.18***	-0.24***
				(0.041)	(0.048)
After School Activities	0.903	0.743	0.681	0.16***	-0.061
				(0.038)	(0.050)
Post HS Exploration	0.329	0.175	0.119	0.15***	-0.057
				(0.041)	(0.039)
Mentorship	1	0.977	0.988	0.023**	0.011
				(0.012)	(0.015)
Extracurricular Activities	0.395	0.175	0.206	0.22***	0.031
				(0.042)	(0.043)
Employment Related Activities	0.264	0.0643	0.0500	0.20***	-0.014
				(0.033)	(0.026)
Contacted Family	0.942	0.971	0.988	-0.029	0.017
				(0.019)	(0.016)
Student Enrichment	0.578	0.310	0.212	0.27***	-0.097**
				(0.047)	(0.048)
Other/Unknown	0.864	0.713	0.663	0.15***	-0.051
				(0.041)	(0.051)
Total Obs.	258	171	160		

Notes: The sample includes only students who enrolled in the CHSS program. Reported means reflect the share of enrolled students who participated in each activity at least once, by district and cohort. Difference columns report simple mean differences between district-cohort groups and are not regression-adjusted. Standard errors in parentheses are heteroskedasticity-robust. *** p<0.01 ** p<0.05 * p<0.10.

Table A.15: ITT Effect of Program on Credit Accumulation, Pooled and by District

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(a) Credits in Year 1			
	(1)	(2)	(3)	(4)
Ever Offer	-0.022	0.018	0.364	-0.303
	(0.111)	–	(0.186)*	(0.234)
	[0.115]	[0.170]	[0.219]*	[0.210]
Control Mean	4.888	5.432	4.380	4.714
N	2318	921	671	726
	(b) Credits Accumulated in Years 1 and 2			
	(1)	(2)	(3)	(4)
Ever Offer	0.028	0.088	0.459	-0.313
	–	(0.229)	(0.382)	(0.234)
	[0.172]	[0.280]	[0.376]	[0.216]
Control Mean	8.368	10.833	8.719	5.510
N	2135	864	644	627
	(c) Above Credit Threshold			
	(1)	(2)	(3)	(4)
Ever Offer	-0.011	0.004	0.027	-0.045
	(0.019)	(0.020)	(0.039)	(0.038)
	[0.020]	[0.029]	[0.039]	[0.035]
Control Mean	0.560	0.613	0.509	0.544
N	2318	921	671	726

Notes: All estimates represent intent-to-treat (ITT) effects, with ever-offer used as the treatment definition and the last complier dropped from each waitlist. All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, spoke English as a Second Language, in poverty, age 15 at enrollment, age 16 or over at enrollment, and whether they have more than two risk factors. The sample is a subset of the 2,360 students in the analytic sample with observable graduation outcomes that also had observable credit accumulation. Outcomes are conditional on enrollment in a school year for at least 10 days and attendance for at least one of those days. Standard errors in parentheses ( ) are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level; those in brackets [ ] are heteroskedasticity-robust. In a small share of cells, the clustered HC2 standard error is left blank because it cannot be reliably computed, as discussed in footnote 12; the heteroskedasticity-robust standard error in brackets is reported in all cases. *** p<0.01 ** p<0.05 * p<0.10.

Table A.16: ITT Effect of Program on Chronic Absenteeism, Pooled and by District

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(a) Chronically Absent in Year 1			
	(1)	(2)	(3)	(4)
Ever Offer	-0.023	-0.030	-0.058	0.014
	(0.015)	(0.022)	(0.023)**	(0.031)
	[0.019]	[0.031]	[0.035]*	[0.032]
Control Mean	0.361	0.463	0.359	0.263
N	2328	925	674	729
	(b) Chronically Absent in Years 1 or 2			
	(1)	(2)	(3)	(4)
Ever Offer	-0.005	-0.029	-0.025	0.031
	–	(0.029)	(0.027)	(0.038)
	[0.019]	[0.029]	[0.037]	[0.035]
Control Mean	0.506	0.616	0.455	0.434
N	2341	931	679	731

Notes: All estimates represent intent-to-treat (ITT) effects, with ever-offer used as the treatment definition and the last complier dropped from each waitlist. All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, spoke English as a Second Language, in poverty, age 15 at enrollment, age 16 or over at enrollment, and whether they have more than two risk factors. The sample is a subset of the 2,360 students in the analytic sample with observable graduation outcomes that also had observable chronic absenteeism. Chronic absenteeism data are unavailable in District B Cohort 17 for Year 3 and District B Cohort 18 for Year 4 due to COVID-19. Both outcomes are conditional on enrollment in a school year for at least 10 days and attendance for at least one of those days. Standard errors in parentheses ( ) are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level; those in brackets [ ] are heteroskedasticity-robust. In a small share of cells, the clustered HC2 standard error is left blank because it cannot be reliably computed, as discussed in footnote 12; the heteroskedasticity-robust standard error in brackets is reported in all cases. *** p<0.01 ** p<0.05 * p<0.10.

Table A.17: ITT Effect of Program on Taking and Passing Graduation Exams in First Year, Pooled and by District

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(a) Took Graduation Exam			
	(1)	(2)	(3)	(4)
Ever Offer	0.022	0.042	0.098	-0.046
	(0.019)	–	(0.032)***	(0.027)
	[0.018]	[0.030]	[0.032]***	[0.031]
Control Mean	0.633	0.392	0.747	0.789
N	2360	941	688	731
	(b) Passed at Least 1 Graduation Exam			
	(1)	(2)	(3)	(4)
Ever Offer	0.016	0.033	0.066	-0.041
	(0.018)	(0.028)	(0.033)*	(0.030)
	[0.019]	[0.029]	[0.037]*	[0.033]
Control Mean	0.388	0.334	0.376	0.449
N	2360	941	688	731

Notes: All estimates represent intent-to-treat (ITT) effects, with ever-offer used as the treatment definition and the last complier dropped from each waitlist. All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, spoke English as a Second Language, in poverty, age 15 at enrollment, age 16 or over at enrollment, and whether they have more than two risk factors. The sample is a subset of the 2,360 students in the analytic sample with observable graduation outcomes that also had observable graduation exam data. Standard errors in parentheses ( ) are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level; those in brackets [ ] are heteroskedasticity-robust. In a small share of cells, the clustered HC2 standard error is left blank because it cannot be reliably computed, as discussed in footnote 12; the heteroskedasticity-robust standard error in brackets is reported in all cases. *** p<0.01 ** p<0.05 * p<0.10.

Table A.18: ITT Effect of Program on High School Dropout, Pooled and by District

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(a) Dropout in Year 1			
	(1)	(2)	(3)	(4)
Ever Offer	-0.002	-0.002	-0.014	0.009
	(0.003)	(0.002)	(0.006)**	(0.007)
	[0.003]	[0.003]	[0.008]*	[0.006]
Control Mean	0.007	0.003	0.017	0.003
N	2354	936	687	731
	(b) Dropout in Year 2			
	(1)	(2)	(3)	(4)
Ever Offer	0.003	-0.000	0.003	0.005
	(0.003)	(0.004)	(0.007)	(0.003)
	[0.003]	[0.004]	[0.008]	[0.004]
Control Mean	0.003	0.003	0.008	0.000
N	2354	936	687	731
	(c) Dropout in Year 3			
	(1)	(2)	(3)	(4)
Ever Offer	0.002	-0.001	0.002	0.001
	(0.005)	(0.009)	(0.011)	(0.008)
	[0.005]	[0.009]	[0.009]	[0.009]
Control Mean	0.016	0.021	0.013	0.012
N	2354	936	687	731
	(d) Dropout in Year 4			
	(1)	(2)	(3)	(4)
Ever Offer	-0.003	-0.002	-0.009	-0.005
	–	–	(0.010)	(0.007)
	[0.007]	[0.016]	[0.010]	[0.010]
Control Mean	0.032	0.055	0.021	0.018
N	2354	936	687	731

Notes: All estimates represent intent-to-treat (ITT) effects, with ever-offer used as the treatment definition and the last complier dropped from each waitlist. All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, spoke English as a Second Language, in poverty, age 15 at enrollment, age 16 or over at enrollment, and whether they have more than two risk factors. The sample is a subset of the 2,360 students in the analytic sample with observable graduation outcomes that also had observable dropout status. Standard errors in parentheses ( ) are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level; those in brackets [ ] are heteroskedasticity-robust. In a small share of cells, the clustered HC2 standard error is left blank because it cannot be reliably computed, as discussed in footnote 12; the heteroskedasticity-robust standard error in brackets is reported in all cases. *** p<0.01 ** p<0.05 * p<0.10.

Table A.19: ITT Effect of Program Early On-Track Indicators: On Track Defined Based on Minimum Credits and Passage of Graduation Exams in Year 1

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(a) Full Sample			
	(1)	(2)	(3)	(4)
Ever Offer	0.015	0.033	0.066	-0.048
	(0.020)	(0.029)	(0.033)*	(0.035)
	[0.018]	[0.028]	[0.036]*	[0.033]
Control Mean	0.376	0.318	0.362	0.439
N	2447	965	711	771
	(b) Analytic Sample			
	(1)	(2)	(3)	(4)
Ever Offer	0.010	0.038	0.051	-0.057
	(0.019)	(0.029)	(0.035)	(0.031)*
	[0.019]	[0.029]	[0.037]	[0.033]*
Control Mean	0.386	0.324	0.386	0.447
N	2318	921	671	726

Notes: All estimates represent intent-to-treat (ITT) effects, with ever-offer used as the treatment definition and the last complier dropped from each waitlist. All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, spoke English as a Second Language, in poverty, age 15 at enrollment, age 16 or over at enrollment, and whether they have more than two risk factors. The on-track indicator equals 1 for students who accumulated at least 2.5 credits in Year 1 and passed at least one required graduation exam in Year 1. Panel (a) reports estimates for the full sample and Panel (b) for the analytic sample; both are restricted to students with non-missing data on all components of the indicator, and students missing any input measure are excluded. Standard errors in parentheses ( ) are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level; those in brackets [ ] are heteroskedasticity-robust. In a small share of cells, the clustered HC2 standard error is left blank because it cannot be reliably computed, as discussed in footnote 12; the heteroskedasticity-robust standard error in brackets is reported in all cases. *** p<0.01 ** p<0.05 * p<0.10.

Table A.20: ITT Effect of Program Early On-Track Indicators: On Track Defined Based on Minimum Credits, Passage of Graduation Exam, Dropout Status, and Absenteeism in Year 1

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(a) Full Sample			
	(1)	(2)	(3)	(4)
Ever Offer	0.017	0.035	0.043	-0.030
	(0.018)	(0.022)	(0.031)	(0.036)
	[0.018]	[0.026]	[0.035]	[0.032]
Control Mean	0.318	0.244	0.325	0.383
N	2435	958	709	768
	(b) Analytic Sample			
	(1)	(2)	(3)	(4)
Ever Offer	0.014	0.041	0.028	-0.036
	–	(0.023)*	(0.032)	(0.036)
	[0.018]	[0.027]	[0.036]	[0.033]
Control Mean	0.326	0.246	0.346	0.389
N	2312	916	670	726

Notes: All estimates represent intent-to-treat (ITT) effects, with ever-offer used as the treatment definition and the last complier dropped from each waitlist. All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, spoke English as a Second Language, in poverty, age 15 at enrollment, age 16 or over at enrollment, and whether they have more than two risk factors. The on-track indicator equals 1 for students who did not drop out in Year 1, accumulated at least 2.5 credits in Year 1, were not chronically absent in Year 1, and passed at least one required graduation exam in Year 1. Panel (a) reports estimates for the full sample and Panel (b) for the analytic sample; both are restricted to students with non-missing data on all components of the indicator, and students missing any input measure are excluded. Standard errors in parentheses ( ) are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level; those in brackets [ ] are heteroskedasticity-robust. In a small share of cells, the clustered HC2 standard error is left blank because it cannot be reliably computed, as discussed in footnote 12; the heteroskedasticity-robust standard error in brackets is reported in all cases. *** p<0.01 ** p<0.05 * p<0.10.

Table A.21: ITT Effect of Program Early On-Track Indicators: On Track Defined Based On Dropout Status, Minimum Credits, and Absenteeism in Year 1

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(a) Full Sample			
	(1)	(2)	(3)	(4)
Ever Offer	-0.002	0.011	0.050	-0.061
	–	(0.021)	(0.031)	(0.046)
	[0.019]	[0.030]	[0.036]	[0.034]*
Control Mean	0.570	0.500	0.568	0.637
N	2435	958	709	768
	(b) Analytic Sample			
	(1)	(2)	(3)	(4)
Ever Offer	-0.000	0.025	0.038	-0.061
	–	(0.021)	(0.032)	(0.046)
	[0.019]	[0.030]	[0.037]	[0.035]*
Control Mean	0.574	0.495	0.588	0.641
N	2312	916	670	726

Notes: All estimates represent intent-to-treat (ITT) effects, with ever-offer used as the treatment definition and the last complier dropped from each waitlist. All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, spoke English as a Second Language, in poverty, age 15 at enrollment, age 16 or over at enrollment, and whether they have more than two risk factors. The on-track indicator equals 1 for students who did not drop out in Year 1, accumulated at least 2.5 credits in Year 1, and were not chronically absent in Year 1. Panel (a) reports estimates for the full sample and Panel (b) for the analytic sample; both are restricted to students with non-missing data on all components of the indicator, and students missing any input measure are excluded. Standard errors in parentheses ( ) are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level; those in brackets [ ] are heteroskedasticity-robust. In a small share of cells, the clustered HC2 standard error is left blank because it cannot be reliably computed, as discussed in footnote 12; the heteroskedasticity-robust standard error in brackets is reported in all cases. *** p<0.01 ** p<0.05 * p<0.10.

Table A.22: ITT Effect of Program on Credit Accumulation, Pooled and by District - Full Sample

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(a) Credits in Year 1			
	(1)	(2)	(3)	(4)
Ever Offer	-0.016	-0.022	0.398	-0.281
	(0.107)	(0.089)	(0.154)**	(0.228)
	[0.113]	[0.167]	[0.215]*	[0.206]
Control Mean	4.866	5.427	4.272	4.749
N	2447	965	711	771
	(b) Credits Accumulated in Years 1 and 2			
	(1)	(2)	(3)	(4)
Ever Offer	0.035	0.067	0.429	-0.274
	–	(0.244)	(0.365)	(0.221)
	[0.171]	[0.276]	[0.375]	[0.212]
Control Mean	8.358	10.853	8.649	5.527
N	2207	893	667	647
	(c) Above Credit Threshold			
	(1)	(2)	(3)	(4)
Ever Offer	-0.012	0.001	0.037	-0.053
	(0.019)	(0.020)	(0.035)	(0.038)
	[0.019]	[0.029]	[0.038]	[0.034]
Control Mean	0.557	0.609	0.490	0.556
N	2447	965	711	771

Notes: All estimates represent intent-to-treat (ITT) effects, with ever-offer used as the treatment definition and the last complier dropped from each waitlist. All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, spoke English as a Second Language, in poverty, age 15 at enrollment, age 16 or over at enrollment, and whether they have more than two risk factors. The sample is a subset of the 2,507 students in the full sample that had observable credit accumulation. Outcomes are conditional on enrollment in a school year for at least 10 days and attendance for at least one of those days. Standard errors in parentheses ( ) are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level; those in brackets [ ] are heteroskedasticity-robust. In a small share of cells, the clustered HC2 standard error is left blank because it cannot be reliably computed, as discussed in footnote 12; the heteroskedasticity-robust standard error in brackets is reported in all cases. *** p<0.01 ** p<0.05 * p<0.10.

Table A.23: ITT Effect of Program on Chronic Absenteeism, Pooled and by District - Full Sample

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(a) Chronically Absent in Year 1			
	(1)	(2)	(3)	(4)
Ever Offer	-0.016	-0.012	-0.068	0.024
	(0.015)	–	(0.026)**	(0.030)
	[0.018]	[0.030]	[0.034]**	[0.031]
Control Mean	0.360	0.456	0.372	0.261
N	2462	973	715	774
	(b) Any Chronic Absenteeism in Years 1 or 2			
	(1)	(2)	(3)	(4)
Ever Offer	-0.001	-0.019	-0.033	0.034
	(0.018)	–	(0.030)	(0.037)
	[0.019]	[0.029]	[0.036]	[0.034]
Control Mean	0.506	0.613	0.470	0.429
N	2477	979	721	777

Notes: All estimates represent intent-to-treat (ITT) effects, with ever-offer used as the treatment definition and the last complier dropped from each waitlist. All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, spoke English as a Second Language, in poverty, age 15 at enrollment, age 16 or over at enrollment, and whether they have more than two risk factors. The sample is a subset of the 2,507 students in the full sample that had observable chronic absenteeism. Chronic absenteeism data are unavailable in District B Cohort 17 for Year 3 and District B Cohort 18 for Year 4 due to COVID-19. Both outcomes are conditional on enrollment in a school year for at least 10 days and attendance for at least one of those days. Standard errors in parentheses ( ) are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level; those in brackets [ ] are heteroskedasticity-robust. In a small share of cells, the clustered HC2 standard error is left blank because it cannot be reliably computed, as discussed in footnote 12; the heteroskedasticity-robust standard error in brackets is reported in all cases. *** p<0.01 ** p<0.05 * p<0.10.

Table A.24: ITT Effect of Program on Taking and Passing Graduation Exams in First Year, Pooled and by District - Full Sample

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(a) Took Graduation Exam			
	(1)	(2)	(3)	(4)
Ever Offer	0.017	0.034	0.093	-0.055
	(0.019)	(0.029)	(0.031)***	(0.029)*
	[0.018]	[0.029]	[0.031]***	[0.030]*
Control Mean	0.626	0.384	0.732	0.786
N	2507	994	734	779
	(b) Passed at Least 1 Graduation Exam			
	(1)	(2)	(3)	(4)
Ever Offer	0.021	0.026	0.080	-0.033
	(0.019)	(0.028)	(0.033)**	(0.033)
	[0.019]	[0.028]	[0.035]**	[0.033]
Control Mean	0.376	0.326	0.354	0.439
N	2507	994	734	779

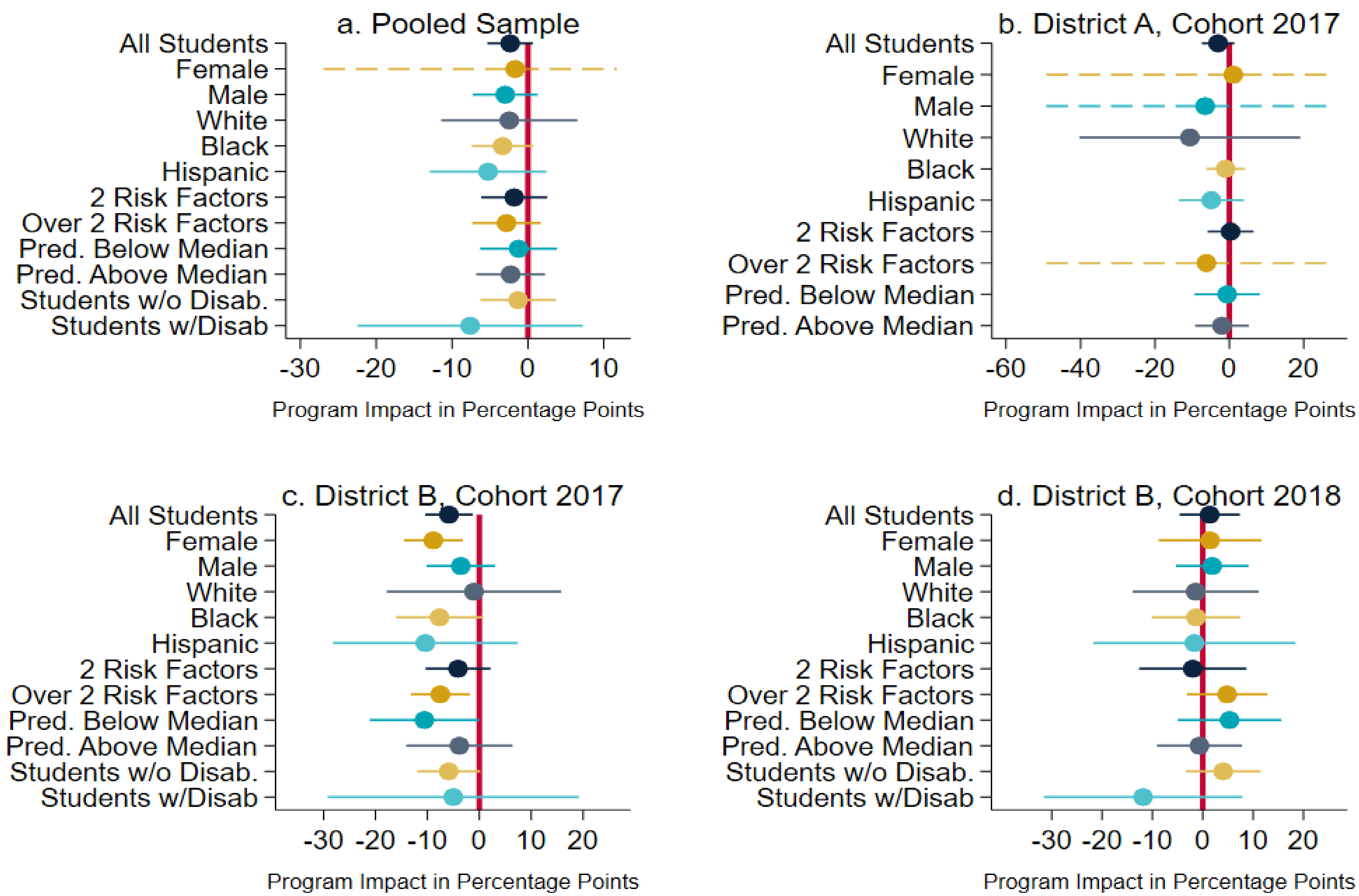
Notes: All estimates represent intent-to-treat (ITT) effects, with ever-offer used as the treatment definition and the last complier dropped from each waitlist. All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, spoke English as a Second Language, in poverty, age 15 at enrollment, age 16 or over at enrollment, and whether they have more than two risk factors. The sample is a subset of the 2,507 students in the full sample that had observable graduation exam data. Standard errors in parentheses ( ) are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level; those in brackets [ ] are heteroskedasticity-robust. *** p<0.01 ** p<0.05 * p<0.10.

Table A.25: ITT Effect of Program on High School Dropout, Pooled and by District - Full Sample

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(a) Dropout in Year 1			
	(1)	(2)	(3)	(4)
Ever Offer	-0.002	-0.002	-0.013	0.008
	(0.003)	(0.002)	(0.006)**	(0.007)
	[0.003]	[0.003]	[0.008]*	[0.006]
Control Mean	0.006	0.003	0.016	0.003
N	2493	987	732	774
	(b) Dropout in Year 2			
	(1)	(2)	(3)	(4)
Ever Offer	0.003	-0.000	0.003	0.005
	(0.003)	(0.004)	(0.006)	(0.003)
	[0.003]	[0.004]	[0.007]	[0.003]
Control Mean	0.003	0.003	0.008	0.000
N	2493	987	732	774
	(c) Dropout in Year 3			
	(1)	(2)	(3)	(4)
Ever Offer	0.001	-0.002	0.001	0.001
	(0.005)	(0.009)	(0.011)	(0.007)
	[0.005]	[0.009]	[0.008]	[0.008]
Control Mean	0.015	0.020	0.012	0.011
N	2493	987	732	774
	(d) Dropout in Year 4			
	(1)	(2)	(3)	(4)
Ever Offer	-0.003	-0.002	-0.008	-0.005
	(0.006)	–	(0.010)	(0.007)
	[0.007]	[0.015]	[0.010]	[0.010]
Control Mean	0.031	0.053	0.020	0.017
N	2493	987	732	774

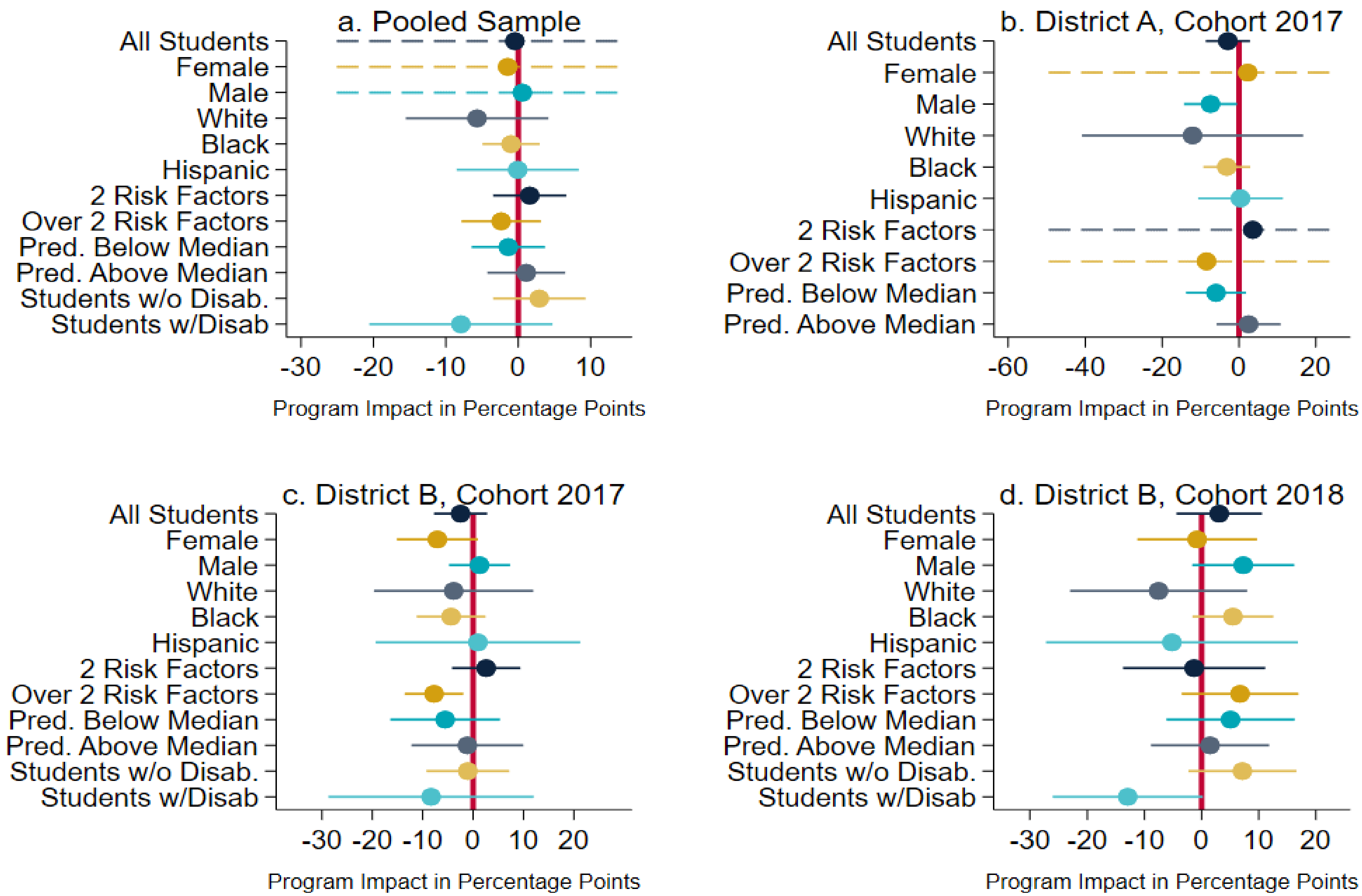
Notes: All estimates represent intent-to-treat (ITT) effects, with ever-offer used as the treatment definition and the last complier dropped from each waitlist. All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, spoke English as a Second Language, in poverty, age 15 at enrollment, age 16 or over at enrollment, and whether they have more than two risk factors. The sample is a subset of the 2,507 students in the full sample that had observable dropout status. Standard errors in parentheses ( ) are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level; those in brackets [ ] are heteroskedasticity-robust. In a small share of cells, the clustered HC2 standard error is left blank because it cannot be reliably computed, as discussed in footnote 12; the heteroskedasticity-robust standard error in brackets is reported in all cases. *** p<0.01 ** p<0.05 * p<0.10

Figure A.1: Intent to Treat Estimates of Effect of Program on Chronically Absent in Year 1



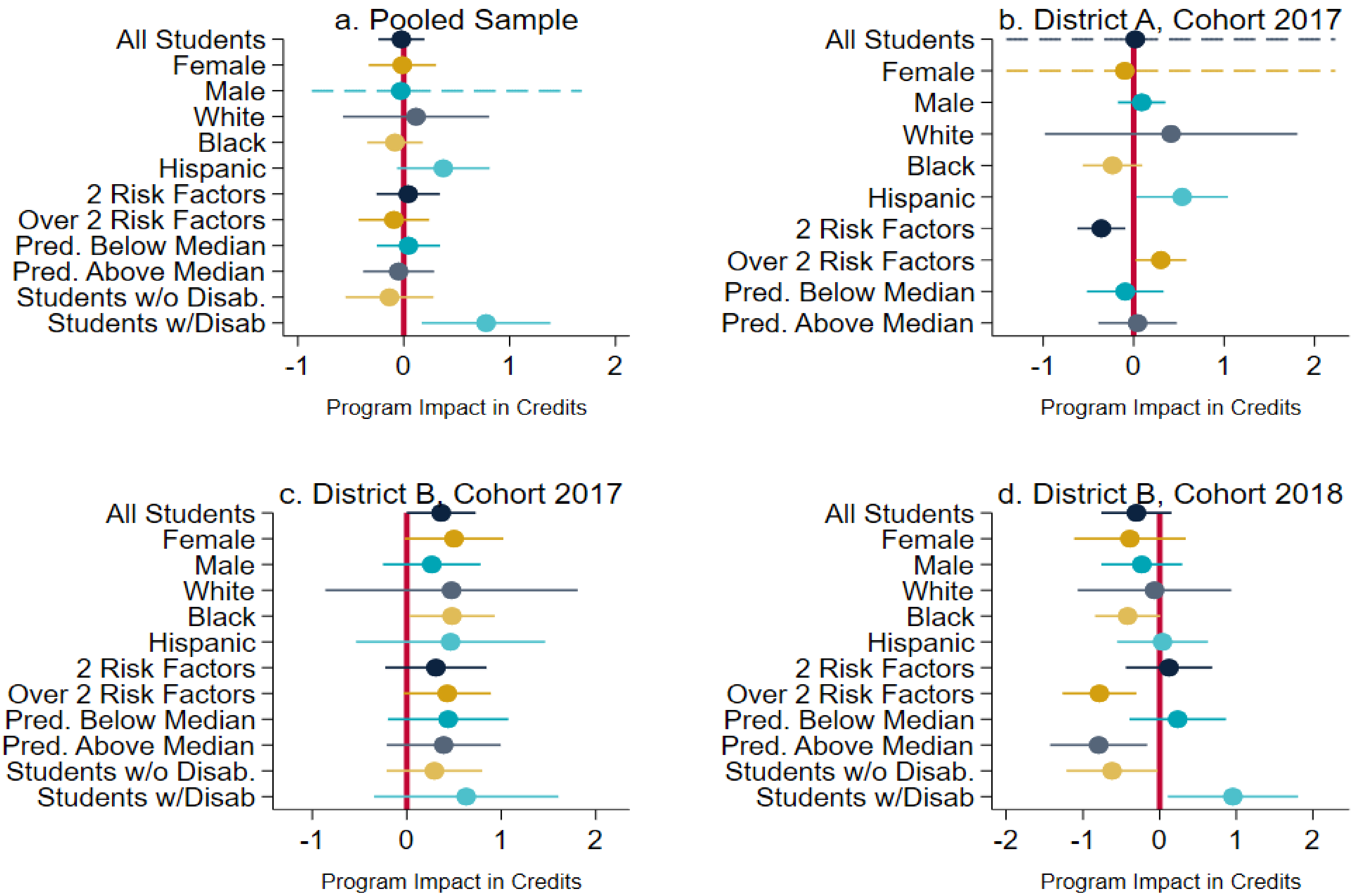
Notes: Each point is the regression-adjusted ITT difference in chronic absenteeism in Year 1 between the treatment and control groups, using the ever-offer treatment definition and dropping the last complier from each waitlist to account for the endogenous stopping rule (de Chaisemartin and Behaghel, 2020). All regressions include school and cohort fixed effects. Standard errors are computed using the reg_sandwich procedure with HC2 bias-corrected inference, clustered at the school/risk-factor-category/gender/cohort level. Bars represent 95% confidence intervals. In a small share of subgroups, the confidence interval cannot be reliably computed and the bars are represented as dashed horizontal lines spanning the figures. Footnote 12 discusses this computation decision. Pred. denotes predicted graduation rate, constructed using baseline covariates following the repeated split-sample procedure of Abadie et al. (2018); the Pred. Below Median" bars represent program impacts for students whose predicted graduation probability falls below the sample median. Students w/o Disab. denotes students without disabilities and Students w/Disab. denotes students with disabilities. District A did not provide information on student disability status.

Figure A.2: Intent to Treat Estimates of Effect of Program on Chronically Absent in Years 1 or 2 for Subgroups



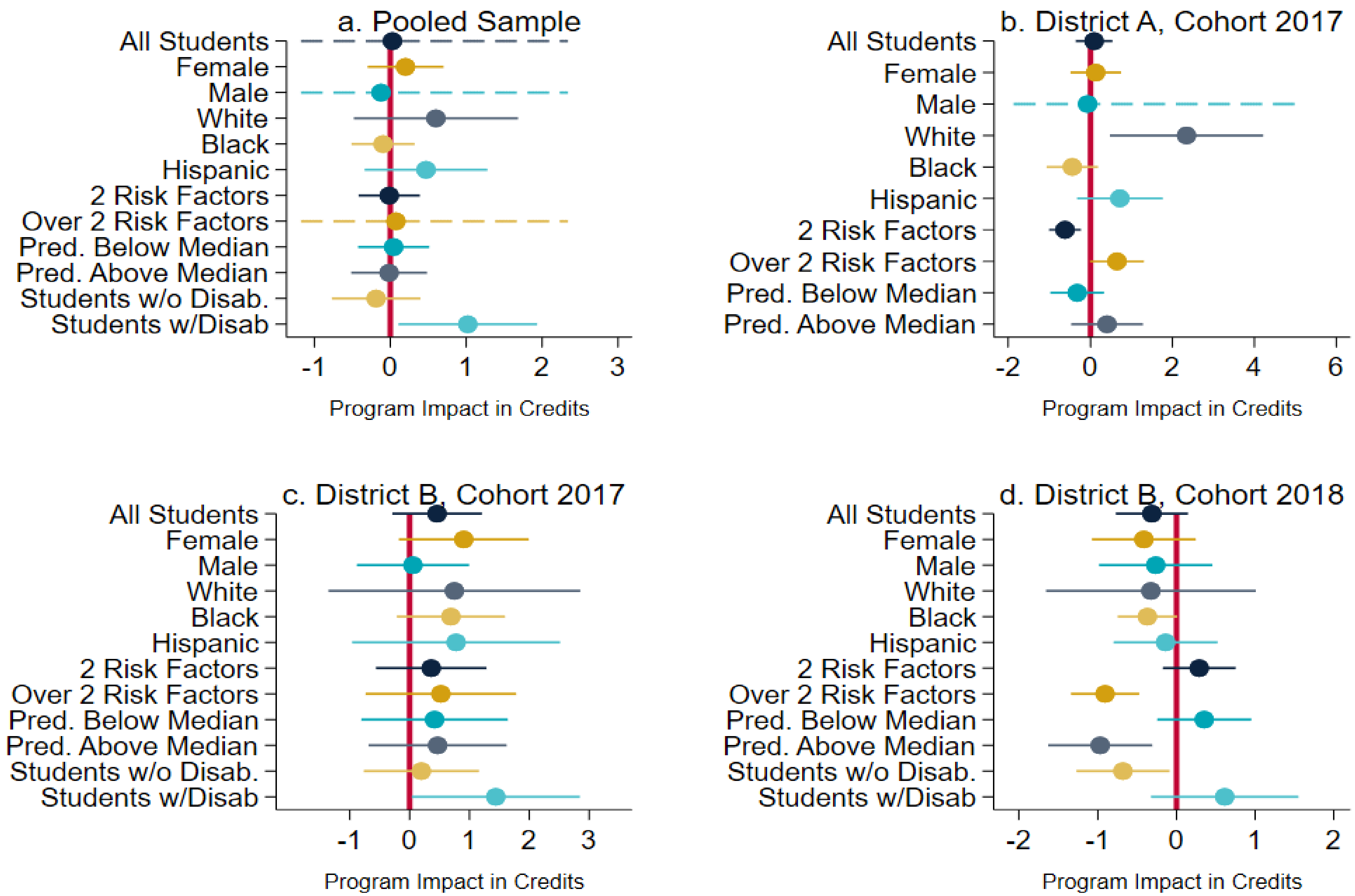
Notes: Each point is the regression-adjusted ITT difference in chronic absenteeism in Years 1 or 2 between the treatment and control groups, using the ever-offer treatment definition and dropping the last complier from each waitlist to account for the endogenous stopping rule (de Chaisemartin and Behaghel, 2020). All regressions include school and cohort fixed effects. Standard errors are computed using the reg_sandwich procedure with HC2 bias-corrected inference, clustered at the school/risk-factor-category/gender/cohort level. Bars represent 95% confidence intervals. In a small share of subgroups, the confidence interval cannot be reliably computed and the bars are represented as dashed horizontal lines spanning the figures. Footnote 12 discusses this computation decision. Pred. denotes predicted graduation rate, constructed using baseline covariates following the repeated split-sample procedure of Abadie et al. (2018); the Pred. Below Median" bars represent program impacts for students whose predicted graduation probability falls below the sample median. Students w/o Disab. denotes students without disabilities and Students w/Disab. denotes students with disabilities. District A did not provide information on student disability status.

Figure A.3: Intent to Treat Estimates of Effect of Program on Credits Achieved in Year 1



Notes: Each point is the regression-adjusted ITT difference in credits achieved in Year 1 between the treatment and control groups, using the ever-offer treatment definition and dropping the last complier from each waitlist to account for the endogenous stopping rule (de Chaisemartin and Behaghel, 2020). All regressions include school and cohort fixed effects. Standard errors are computed using the reg_sandwich procedure with HC2 bias-corrected inference, clustered at the school/risk-factor-category/gender/cohort level. Bars represent 95% confidence intervals. In a small share of subgroups, the confidence interval cannot be reliably computed and the bars are represented as dashed horizontal lines spanning the figures. Footnote 12 discusses this computation decision. Pred. denotes predicted graduation rate, constructed using baseline covariates following the repeated split-sample procedure of Abadie et al. (2018); the Pred. Below Median" bars represent program impacts for students whose predicted graduation probability falls below the sample median. Students w/o Disab. denotes students without disabilities and Students w/Disab. denotes students with disabilities. District A did not provide information on student disability status.

Figure A.4: Intent to Treat Estimates of Effect of Program on Cumulative Credits Achieved in Years 1 and 2



Notes: Each point is the regression-adjusted ITT difference in credits achieved in Year 1 between the treatment and control groups, using the ever-offer treatment definition and dropping the last complier from each waitlist to account for the endogenous stopping rule (de Chaisemartin and Behaghel, 2020). All regressions include school and cohort fixed effects. Standard errors are computed using the reg_sandwich procedure with HC2 bias-corrected inference, clustered at the school/risk-factor-category/gender/cohort level. Bars represent 95% confidence intervals. In a small share of subgroups, the confidence interval cannot be reliably computed and the bars are represented as dashed horizontal lines spanning the figures. Footnote 12 discusses this computation decision. Pred. denotes predicted graduation rate, constructed using baseline covariates following the repeated split-sample procedure of Abadie et al. (2018); the Pred. Below Median" bars represent program impacts for students whose predicted graduation probability falls below the sample median. Students w/o Disab. denotes students without disabilities and Students w/Disab. denotes students with disabilities. District A did not provide information on student disability status.

B Appendix B

2.1 Post-Secondary Outcomes

As an exploratory analysis of longer-run impacts, we examine program effects on post-secondary enrollment and degree attainment using data from the National Student Clearinghouse.[21,22] We find no evidence that CHSS increased the likelihood of ever enrolling in college or enrolling within one year of high school graduation for the pooled sample or for any subgroup (see Appendix Table B.1).[23] In fact, estimates for the pooled sample, and for District B in particular, are negative and statistically significant, driven primarily by the lower high school graduation rates observed for the 2018 cohort in District B. Conditional on enrolling in post-secondary education, we find no significant difference in total semesters enrolled between treatment and control students. Finally, we find little evidence of an effect on post-secondary degree completion for the pooled sample. We should caution that because we only have data for either seven or eight years after initial enrollment in the program, only students who enrolled soon after graduating high school and who maintained a fairly full course load would have the opportunity to complete a two-year post-secondary degree within this period. This is evident in the means for the control group — less than 2.5 percent of the control group in the pooled sample has completed a degree within this time frame. Our estimates suggest that the program had little effect on degree completion for the pooled sample. There is some evidence that the program increased post-secondary graduation in District A. Those in the treatment group were 1.7 percentage points more likely to have graduated, and this effect is marginally significant.

[21] The National Student Clearinghouse is a nonprofit organization founded in 1993 that provides nationwide verification of student enrollment and degree attainment and currently covers institutions enrolling approximately 97 percent of U.S. college students.

[22] Opportunities related to both college access and career exploration were substantially constrained following the onset of the COVID-19 pandemic, including limitations on in-person advising, college visits, internships, and entry-level employment opportunities. These disruptions may have attenuated post-secondary responses to the program, particularly for cohorts whose high school experience coincided with the pandemic.

[23] Results are similar when estimated using the full sample; see Appendix Table B.2.

Table B.1: Effect of Program on Post-Secondary Outcomes, Pooled and by District - Analytic Sample

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(a) Enrolled in Post-Secondary Education as of December 2024			
	(1)	(2)	(3)	(4)
Ever Offer	-0.028	-0.016	-0.016	-0.053
	(0.011)**	(0.016)	(0.025)	(0.017)***
	[0.014]**	[0.019]	[0.026]	[0.027]**
Control Mean	0.144	0.098	0.131	0.199
N	2349	930	688	731
	(b) Total Semesters Enrolled as of December 2024			
	(1)	(2)	(3)	(4)
Ever Offer	0.061	0.014	0.044	0.032
	(0.038)	(0.096)	(0.057)	(0.065)
	[0.043]	[0.067]	[0.079]	[0.075]
Control Mean	1.085	1.094	1.097	1.076
N	282	83	81	118
	(c) Completed Post-Secondary Education as of December 2024			
	(1)	(2)	(3)	(4)
Ever Offer	0.011	0.017	0.015	0.001
	(0.006)*	(0.010)	(0.013)	(0.009)
	[0.007]	[0.009]*	[0.016]	[0.013]
Control Mean	0.025	0.012	0.034	0.030
N	2349	930	688	731

Notes: All estimates represent intent-to-treat (ITT) effects, with ever-offer used as the treatment definition and the last complier dropped from each waitlist. All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, spoke English as a Second Language, in poverty, age 15 at enrollment, age 16 or over at enrollment, and whether they have more than two risk factors. The sample includes all students in the analytic sample who also matched to National Student Clearinghouse (NSC) records, which requires non-missing first name, last name, and date of birth; NSC data are available through December 31, 2024. Total semesters enrolled is conditional on having enrolled in post-secondary education. Standard errors in parentheses ( ) are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level; those in brackets [ ] are heteroskedasticity-robust. *** p<0.01 ** p<0.05 * p<0.10.

Table B.2: Effect of Program on Post-Secondary Outcomes, Pooled and by District - Full Sample

	Pooled	District A	District B	
	C17 & C18	C17	C17	C18
	(a) Enrolled in Post-Secondary Education as of December 2024			
	(1)	(2)	(3)	(4)
Ever Offer	-0.024	-0.020	-0.002	-0.045
	(0.011)**	(0.015)	(0.024)	(0.019)**
	[0.013]*	[0.019]	[0.025]	[0.026]*
Control Mean	0.142	0.101	0.122	0.194
N	2480	968	734	778
	(b) Total Semesters Enrolled as of December 2024			
	(1)	(2)	(3)	(4)
Ever Offer	0.070	0.033	0.066	0.047
	(0.038)*	(0.084)	(0.059)	(0.062)
	[0.043]	[0.064]	[0.084]	[0.074]
Control Mean	1.082	1.088	1.097	1.072
N	298	85	87	126
	(c) Completed Post-Secondary Education as of December 2024			
	(1)	(2)	(3)	(4)
Ever Offer	0.013	0.016	0.018	0.003
	(0.006)**	(0.009)	(0.013)	(0.009)
	[0.007]*	[0.010]*	[0.015]	[0.013]
Control Mean	0.024	0.015	0.031	0.028
N	2480	968	734	778

Notes: All estimates represent intent-to-treat (ITT) effects, with ever-offer used as the treatment definition and the last complier dropped from each waitlist. All regressions include school and cohort fixed effects. Control variables include binary indicators for whether a student is female, Black, White, Hispanic, spoke English as a Second Language, in poverty, age 15 at enrollment, age 16 or over at enrollment, and whether they have more than two risk factors. The sample includes all students in the full sample who also matched to National Student Clearinghouse (NSC) records, which requires non-missing first name, last name, and date of birth; NSC data are available through December 31, 2024. Total semesters enrolled is conditional on having enrolled in post-secondary education. Standard errors in parentheses ( ) are bias-corrected sandwich (HC2) standard errors clustered at the cohort/school/risk factor category/gender level; those in brackets [ ] are heteroskedasticity-robust. *** p<0.01 ** p<0.05 * p<0.10.