When Resources Meet Relationships: The Returns to Personalized Supports for Low-Income Students*

Benjamin Goldman, Cornell University
Jamie Gracie, Harvard University†

December 2025

Abstract

Children from low-income families face persistent educational and economic disadvantages. This paper studies Communities In Schools (CIS), a program that places coordinators in high-poverty schools to connect struggling students with personalized support. CIS is the largest program of its kind in the US, reaching 2 million students each year—nearly three times the size of Head Start—and, unlike most programs of this scale, is funded largely by private philanthropy and local governments. Using the staggered rollout of CIS, we find that the program boosts test scores for struggling students, and that these improvements persist, ultimately increasing high-school completion and adult earnings. These long-run effects can be closely forecast from changes in short-run outcomes, with non-cognitive measures playing a central role. CIS emphasizes personalization as a core feature of its model, and our results are consistent with this claim: coordinators tailor services to distinct student needs, yet students with different needs still see similarly large long-run gains. CIS delivers returns that compare favorably to other major education interventions, such as class-size reductions.

*The conclusions of this research do not necessarily reflect the opinion or official position of the Texas Education Research Center, the Texas Education Agency, the Texas Higher Education Coordinating Board, the Texas Workforce Commission, or the State of Texas. Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau. The Census Bureau has ensured appropriate access and use of confidential data and has reviewed these results for disclosure avoidance protection (Project 751987: CBDRB-FY23-CES014-028). We thank Raj Chetty, Sahil Chinoy, Heather Clawson, John Friedman, Claudia Goldin, Nathan Hendren, Larry Katz, Martin Koenen, Kevin Leary, Matthew Notowodigdo, Amanda Pallais, Lynne Sacks, and Rob Watson for helpful feedback. This research was funded by Opportunity Insights, The EdRedesign Lab at Harvard Graduate School of Education, the Stone Program at Harvard Kennedy School, the National Science Foundation (Graduate Research Fellowship, grant DGE-2140743), and Cornell University.

†Corresponding author: jgracie@g.harvard.edu

I Introduction

In the United States, individuals from low-income families earn significantly less than their peers from high-income families (Chetty et al., 2014; Solon, 1992). These disparities begin well before labor market entry: children from low-income backgrounds score lower on standardized tests, are less likely to graduate from high school, and attend college at lower rates (Black, Devereux and Salvanes, 2005; Chetty et al., 2020). Improving educational outcomes for children from low-income families is likely to be a promising way to reduce earnings disparities later in life, given the strong evidence on the returns to education (Bhuller, Mogstad and Salvanes, 2017; Card, 1999; Zimmerman, 2014).

One approach is to increase resources and programs in high-poverty schools. Initiatives to improve teacher quality, reduce class sizes, expand afterschool and tutoring programs, and equalize school funding have all shown important benefits (Aizer et al., 2016; Chetty, Friedman and Rockoff, 2014; Hyman, 2017; Jackson, Johnson and Persico, 2016; Nickow, Oreopoulos and Quan, 2020). Still, many low-income students face barriers that hinder their academic progress: they are more likely to miss school, face logistical or informational hurdles to participating in enrichment programs, and contend with unstable home environments that disrupt learning (Malkus, 2024; Robinson, Bisht and Loeb, 2025). Interventions that provide more comprehensive student support—by addressing academic, social, and non-academic barriers together—may therefore improve outcomes both by complementing existing school resources and by directly meeting students' needs in contexts where such resources remain limited.

How can we design interventions that help students overcome these upstream barriers to learning in high-poverty schools? Evidence from recent randomized trials in other domains points to a promising answer: pairing traditional resources with personalized supports that help individuals navigate challenges and connect to services (e.g., Bergman et al., 2024; Weiss et al., 2019). These interventions rely on the presence of a trusted adult who can coordinate services and advocate for participants. While most existing implementations have focused on adults outside of schools, extending this approach to children in educational settings may both improve outcomes directly and amplify the returns to other school-based investments.

In this paper, we study Communities In Schools (CIS), a program that places a "site coordinator" in high-poverty schools to connect struggling students with both in-school supports (such as tutoring and mentoring) and out-of-school resources (including food assistance, housing support, and health services), tailored to their individual needs. CIS is the largest integrated student support program in the United States, serving nearly 2 million students across more than 3,000 schools each year. This scale is unusual for a privately run program funded largely by private philanthropy, school districts, and local governments rather than federal appropriations

(CIS of Houston, 2024). For context, it is nearly three times the size of Head Start (Office of Head Start, 2025). The goal is to ensure students' basic needs are met so they can succeed in the classroom and beyond. We evaluate whether the program improves students' short-run academic outcomes—test scores, absences, suspensions—and long-run outcomes—high school graduation, college attendance, and earnings—using administrative data from the universe of federal tax records and from K–12 education records in Texas, where nearly half of CIS schools are located.[1]

Our analysis yields five main findings. First, using the staggered rollout of CIS in Texas, we show that the program raises test scores among students who struggle academically, have poor attendance, or face disciplinary problems. Second, these gains do not fade out: they translate into higher rates of high school graduation, two-year college enrollment, and ultimately higher adult earnings, as shown in linked Census and IRS records. Third, the long-run effects of CIS can be predicted from its short-run impacts on test scores, attendance, and discipline, with little evidence of additional channels. Non-cognitive outcomes such as attendance and discipline are central to this result: only about half of the long-run impact on graduation is explained by short-run improvements in test scores. Fourth, the data suggest that the personalization central to the CIS model is efficient—students with different needs receive different services but experience similar long-run gains. Finally, we show that CIS delivers gains per dollar spent that compare favorably to reducing class size, and that targeted expansions could meaningfully raise high school graduation rates among at-risk students. Together, these results show that providing individualized supports in high-poverty schools can both improve student achievement in the short run and translate into durable gains in education and earnings.

CIS operates through a national network of local nonprofit "affiliates" that partner with districts and schools to implement a common model of integrated student supports. New affiliates form through a structured process in which CIS identifies local leaders, convenes a task force, assesses community needs, and develops a work plan before incorporating a new nonprofit with its own board and executive director (CIS, 2025*d*; Attea and Tuck, 2005). Once established, affiliates embed full-time CIS site coordinators in partner schools—typically low-income schools with low student achievement (CIS, 2025*b*; MDRC, 2025; CIS, 2025*c*). In our Texas setting, CIS schools are high-poverty and racially diverse and located in neighborhoods with persistently low adult educational attainment and low upward mobility. Conversations with affiliate directors indicate that school entry is driven mainly by persistent perceptions of need and the presence of a willing school partner; timing is shaped by funding, staffing, and other logistical constraints rather than year-on-year fluctuations in outcomes or changes in school leadership.

[1]Texas is unique in providing substantial state funding for CIS. In 2022, the state allocated more than $38 million to CIS programs, drawing on both the state budget and federal Temporary Assistance for Needy Families funds (Texas Education Agency, 2022).

Within a CIS school, all students can interact with the site coordinator, but the most intensive support is delivered through "case management." Teachers, counselors, and administrators refer students with academic, behavioral, attendance, basic-needs, or social-emotional concerns, and case-managed students meet regularly with the coordinator to ensure they have the resources needed to succeed. In our Texas sample, 12.5% of students receive case management in a given year. The median case-managed student meets with the coordinator 29 times—nearly once per week during the school year. Coordinators use these meetings to connect students to a wide set of supports, including tutoring, test preparation, and help with study habits and course planning. To address non-academic needs, coordinators provide or broker access to social-emotional supports (small-group counseling, goal-setting, mentoring), basic-needs assistance (food pantries, transportation, clothing, emergency housing referrals), and health or mental-health services (Parise and Hossain, 2017; McShane, 2019; CIS, 2025*a*). Services are delivered directly by coordinators, by school-based staff, or through public programs and community partners, with the mix varying across affiliates. Coordinators also run school-wide initiatives—such as attendance campaigns, climate-building activities, and family-engagement events—that aim to strengthen attendance culture and the broader learning environment for all students.

We begin by showing that case management assignments can be predicted with a relatively high degree of accuracy using data from before the CIS program begins. This allows us to construct a natural control group of students in non-CIS schools who would likely have received case management, while avoiding bias from conditioning on case management itself (e.g., a parental divorce may trigger case management and independently lower test scores). Our baseline estimates focus on Texas middle schools with CIS programs, both because standardized test coverage is substantially higher in middle school and because we use elementary school test scores, attendance, suspensions, and student demographics (e.g., free and reduced-price lunch, or FRPL status) to predict case management assignments for middle-school students. A gradient boosting algorithm produces strong out-of-sample predictions: 27.3% of students in the top quartile of predicted case management actually receive services, compared to 3.6% in the bottom quartile. The prediction is even sharper in the tails: 37.4% of students in the top ventile receive services, compared to just 1.1% in the bottom ventile. Likely case managed students are more likely to be FRPL-eligible, have disciplinary records, be chronically absent, and score lower on standardized tests before CIS exposure. Models restricted to a single variable, such as test scores, perform substantially worse. Throughout, we refer to students in the top and bottom quartiles of predicted case management as *high-risk* and *low-risk*, reflecting case management's stated goal of serving students at risk of dropping out in the future.

We next show that CIS improves standardized math and reading scores for high-risk students. In our baseline specification, we estimate an event study where each CIS school is matched to

a "never treated" control school that is demographically and academically similar. The match is constructed using data from more than five years before CIS begins, which allows us to show that outcomes in treated and control schools follow parallel trends prior to CIS entry. We estimate that three years of exposure to CIS increases math scores for high-risk students—relative to their high-risk counterparts in control schools—by 0.18 standard deviations (SD; standard error = 0.04). For low-risk students, we find relatively precise null effects, indicating that CIS's academic benefits are concentrated among students with prior academic, attendance, or discipline challenges. We do not interpret this heterogeneity as arising solely from the causal effect of case management, which would require an additional exclusion restriction. Instead, we view it as evidence that CIS is most effective for students who were struggling in elementary school and less so for those already on track.

We use several strategies to test the robustness of our baseline findings on academic outcomes. Results are similar when we draw comparison groups from outside the matched control, including "later treated" CIS schools. We also relax the parallel trends assumption by allowing trends to vary at the school by race by sex level. As another check, we construct a control group of schools within the same commuting zone as the CIS school, ensuring that our estimates are not driven by city-wide changes coinciding with CIS entry. We find no evidence of changes in student demographics, teacher quality, or teacher tenure around the time CIS enters a school. We also show that results are robust to assigning children to a middle school based on their first elementary school, ruling out the possibility that higher-ability students sort into CIS schools after program entry. In addition, we show that within CIS schools, test score gaps between high- and low-risk students narrow immediately after the program begins, ruling out the possibility that our results reflect school-wide improvements coinciding with CIS entry.

Having established that CIS improves test scores for high-risk students, we next turn to its impact on long-run outcomes: high school graduation, college attendance, and earnings. The mission of CIS is not only to raise in-school performance but also to provide students with the resources they need to succeed beyond school. As the organization itself describes, its goal is to "surround students with a community of support, empowering them to stay in school and achieve in life" (CIS, 2011).

We find that CIS improves long-run outcomes for high-risk students. In our baseline matched-control design, three years of CIS exposure in middle school raises high school graduation by 3.4pp (5.2%; S.E.=1.2pp), closing 13% of the initial gap between high- and low-risk students. Although estimated imprecisely, we also find a 2.9pp (9.1%; S.E.=1.8pp) increase in college attendance, driven entirely by two-year college enrollment—the most common postsecondary path among high-risk students who attend college. For low-risk students, we find relatively precise null effects, consistent with our short-run results.

We complement our analysis of academic outcomes in Texas with linked Census and IRS records to estimate the effect of CIS on earnings (Goldman, Gracie and Porter, 2023). Because most Texas programs began after 2010, many exposed students are still too young to have reliable earnings measures. To address this, we use the staggered rollout of CIS programs outside Texas to estimate event studies and find that three years of exposure raises earnings at age 27 by 1.54 percentile ranks (4.3%; S.E.=0.5pp), corresponding to approximately $\$1,140$. Since we cannot distinguish high- and low-risk students in this sample, this estimate reflects the average effect across the full school population. In Texas, where we measure earnings at an earlier age, we find similarly sized effects for the full population. The gains come mainly from reducing the likelihood that students fall into the bottom quartile of the income distribution and increasing the likelihood that they move into the second quartile, with no effect on the probability of reaching the upper end of the income distribution. Consistent with CIS's description as the nation's "leading dropout prevention organization," the evidence suggests that CIS helps students avoid adverse, or "left-tail," outcomes both in school and later in life.

In the final section, we explore *how* CIS achieves its long-run impacts by linking short-run improvements to long-run outcomes and examining whether its personalized service model contributes to these gains. We first test whether CIS's long-run effects can be anticipated from its short-run impacts on in-school outcomes—specifically, standardized test scores, attendance, and disciplinary incidents. This exercise reveals which dimensions of student progress are most predictive of lasting gains. We find that CIS's long-run effects can be accurately forecast from these short-run measures: CIS raises high school graduation by 3.1pp (SE=1.0pp), while the predicted effect based only on short-run data is 2.4pp (SE=0.6pp). The prediction using short-run measures is accurate for both high- and low-risk students, indicating that standard in-school metrics provide a reliable window into CIS's long-run effectiveness. Notably, if we had relied only on test scores, we would have substantially understated CIS's true impact. Beyond cognitive skills, CIS's lasting effects on engagement and behavior are central to improvements in long-run outcomes.

We next turn to the role of *personalization*, a cornerstone of the CIS model. Site coordinators tailor services to each student's specific needs, which could generate either efficient or inefficient variation in services across students. If the variation reflects misallocation, some students may receive suboptimal bundles and experience smaller gains. If, instead, coordinators match services efficiently to students' needs, variation in service portfolios should align with differences in underlying challenges but yield similar overall gains. To test this, we classify high-risk students by whether their primary needs are academic or non-cognitive. Coordinators indeed provide distinct service portfolios to these groups: students with academic needs receive more academic and cognitive support and show the largest short-run test score gains. Yet the long-run effects of CIS are nearly identical for both groups, suggesting that the data are consistent with efficient allocation,

in which coordinators adapt their approach to individual needs while achieving comparable improvements in long-term outcomes.

Our paper contributes to the literature on improving outcomes in high-poverty schools by evaluating a coordinator-led model that connects struggling students to academic and non-academic supports. Prior studies have largely focused on expanding resources, improving curricula, or introducing new programs—examples include high-dosage tutoring, teacher quality initiatives, behavior interventions, class-size reductions, "no excuses" charter schools, and school finance reforms (e.g., Alan and Kubilay, 2025; Alan, Boneva and Ertac, 2019; Guryan et al., 2023; Chetty, Friedman and Rockoff, 2014; Angrist et al., 2019; Dobbie and Fryer Jr, 2011; Jackson, Johnson and Persico, 2016). We provide large-scale evidence that a comprehensive individualized support model—CIS—raises test scores, attendance, and behavior for high-risk students, with gains that persist through high school graduation, college enrollment, and early-career earnings.[2] On a cost-adjusted basis, CIS compares favorably to other major interventions: each $1,000 of upfront investment raises earnings by $406, compared with roughly $40 for class-size reductions (Chetty et al., 2011).[3] Taken together with evidence from similar interventions in other settings (Lavecchia, Oreopoulos and Brown, 2020), our results show that integrated student supports can deliver durable gains for poor students.

This study also contributes to a growing literature on the role of social connections in shaping economic opportunity. Social networks are an important determinant of economic mobility (Chetty et al., 2022; Fischer and Mattson, 2009; Putnam, 2016), shape how local shocks persist across generations (Chetty et al., 2025*b*), and causally influence human capital investments (Cattan, Salvanes and Tominey, 2025). A distinctive feature of CIS is its reliance on site coordinators—caring adults who work directly with students to identify needs, connect them to supports, and ensure they remain engaged in school. This relational model produces lasting gains in academic achievement, educational attainment, and earnings. Our findings add to a growing body of experimental evidence showing that social ties not only promote mobility directly but also amplify the impact of traditional opportunity-enhancing investments—such as college access

[2]Prior work on CIS is largely descriptive and school-level, with limited counterfactuals and narrow coverage of short-run outcomes (e.g., Porowski and Passa, 2011; Somers and Haider, 2017). MDRC also ran an RCT of case management in 24 schools in North Carolina and Texas with 1,501 students, finding largely null and somewhat imprecise academic effects (Corrin et al., 2015; Parise et al., 2017). Because control students were in the same schools, the RCT nets out any school-wide or spillover effects. It also tracks outcomes for only two years, whereas our event studies show effects that build more gradually. We instead use large-scale student-level administrative data to estimate the causal effect of attending a CIS school on short- and long-run outcomes by comparing likely-affected students to similar peers in fully untreated schools.

[3]We cannot compute a full cost–benefit ratio or marginal value of public funds from the government's perspective because we lack cost data on the services to which CIS students are referred and causal estimates of CIS-induced take-up of public programs. Nonetheless, our earnings estimates imply an increase in the present discounted value of tax revenue of $\$7,100$ per student receiving three years of CIS.

programs, emergency rent assistance, housing vouchers, and job training—when those investments are paired with mentorship and relational support (Barr and Castleman, 2025; Bell, 2025; Bergman et al., 2024; Caspi and Rafkin, 2025; Espinosa et al., 2024; Finkelstein and Notowidigdo, 2019; Katz et al., 2022; Weiss et al., 2019).[4] Unlike most prior studies, however, this paper examines these dynamics in the context of children and K–12 schools. This focus underscores that social connections are not only powerful for adults navigating the labor market and public programs, but can also be deliberately cultivated early in life to alter life trajectories from an early age.

Finally, we contribute to a growing literature that seeks to forecast the long-run returns of interventions focused on children using short- or medium-run academic outcomes. A common practice for school-based programs is to estimate returns on investment by projecting long-run effects from test score gains (e.g., Guryan et al., 2023; Hanushek, 2011; Kline and Walters, 2016; You, Fu and Chou, 2025).[5] We show that CIS's long-run impacts can indeed be forecast from short-run outcomes, but not from test scores alone. Measures of engagement and behavior are essential to capture the noncognitive dimensions through which CIS operates. This finding reinforces evidence that test scores can understate the lasting impacts of programs, especially those that build noncognitive skills (e.g., Anders, Barr and Smith, 2023; Chetty et al., 2011; Cortes, Goodman and Nomi, 2015; Gray-Lobe, Pathak and Walters, 2023). More broadly, our results underscore the promise of using richer sets of short-run outcomes as statistical surrogates to evaluate programs in real time, without waiting years or decades for long-run follow-up data (Athey et al., 2025; Prentice, 1989).

II Background on Communities In Schools

Communities In Schools (CIS) is the largest and most established provider of integrated student support (ISS) services in the United States. Founded in the 1970s in New York, today the program operates in more than 3,500 schools and serves 2 million students across 29 states and the District of Columbia (CIS, 2023). CIS aims to reduce dropout rates and improve academic outcomes by addressing both academic and non-academic barriers to learning. The key component of the CIS model is the "site coordinator"–or navigator–placed in each school. Site coordinators work with school staff to identify students in need, assess the root causes of disengagement, and connect them to tailored supports. Importantly, the site-coordinator is a dedicated CIS employee, not a teacher or administrator, freeing up those members of the school to focus on their own roles.

Integrated student support programs, such as CIS, recognize that many students—especially

[4] One exception is among formerly incarcerated individuals, where similar case-management approaches after release show little impact on post-release outcomes (Doleac, 2019).

[5] The scaling factor from Chetty, Friedman and Rockoff (2014) is frequently used for this purpose. See Table A.IV in Kline and Walters (2016) for a review of observational and experimental estimates linking test scores to earnings.

those from disadvantaged backgrounds—face personal, familial, and community-level challenges that interfere with their ability to succeed in school, including food insecurity, unstable housing, mental health needs, and limited academic support outside the classroom. The navigator's job is to systematically identify these barriers for a given student, connect the student to the necessary set of resources, and ensure that the student is able to fully engage with their education. In addition to providing some direct support—such as academic encouragement and mentoring—navigators serve as bridges linking students to existing school, community, and government services.

Although CIS is a national program, its presence is especially pronounced in Texas, where nearly half of all CIS schools are located. Unlike in other states, CIS is written into the state education charter and receives dedicated public funding both from the Texas Education Agency (TEA) and through federal Temporary Assistance for Needy Families (TANF) funds (Texas Education Agency, 2022). In FY 2022, state funding for CIS totaled $33.5 million plus $4.9 million of TANF funds. Figure I shows the share of students enrolled in a CIS school over time. The program expanded rapidly in recent decades: in 1994, only 1.6% of Texas public school students attended a school with CIS, compared to 12.7% by 2019.

When CIS enters a school, all students become eligible to work with the site coordinator. At the most targeted level, students become enrolled in "case management", which involves regular check-ins with the coordinator. There is no strict rule for how students are referred for case management. In practice, students are often referred based on observable risk factors such as chronic absenteeism, behavioral issues, low academic performance, or signs of unmet basic needs (Parise and Hossain, 2017).[6] In many cases, teachers, counselors, or administrator may flag students for referral. In others, coordinators may proactively identify students using administrative records.

Once students are referred to the site coordinator, they undergo an individualized needs assessment to identify the academic, behavioral, and non-academic barriers that may be impeding their success. The coordinator then works with the student, their family, and school personnel to develop a personalized support plan and begin connecting the student to appropriate services. These supports span a wide range. Academically, students might receive tutoring, credit recovery, test prep, or help with study habits. Outside the classroom, coordinators can provide or facilitate access to food pantries, transportation assistance, medical or mental health care, crisis housing, or family support services. Navigators may also offer mentoring, lead social-emotional learning groups, or organize goal-setting and attendance check-ins. Supports may be delivered directly, through school-based staff, or via external providers and community partnerships. Site

[6]Because CIS receives substantial public funding in Texas, the state imposes additional eligibility guidelines. Students may be enrolled in case management only if they are classified as "at-risk" by TEA, eligible for free or reduced-price lunch, engaging in delinquent activities, or experiencing family conflict or crisis (Texas Education Agency, 2021). In practice, however, these designations are widespread in CIS schools.

coordinators are trained to monitor progress, maintain regular contact with students, and revise plans as needs evolve—serving as a single point of coordination in an often fragmented support ecosystem. More than just a service broker, the navigator also provides students with a consistent, supportive relationship as a "caring adult," a critical element for many students navigating instability at home or in school.

Case management is a high-touch intervention: the median case-managed student meets with the coordinator 29 times per year—nearly once per week during the school year. In addition to formal case management, CIS also implements school-wide initiatives such as attendance drives, behavioral support campaigns, or family engagement events intended to improve school climate and support all students.

Given the goals of the program, it is not surprising that CIS tends to operate in schools and neighborhoods with high poverty rates. Figure II Panel *A*, maps the locations of CIS schools in Texas as of 2019, with counties shaded by their poverty rates in the year 2000. CIS programs are heavily concentrated in the western and southern regions of the state—areas with the highest poverty levels—and are also clustered in major urban centers in the east.

County-level poverty rates do not tell the full story. In Figure II Panel *B*, we compare the characteristics of neighborhoods directly served by CIS schools to those without CIS, using school catchment zones to map neighborhoods to schools. Students in CIS neighborhoods are substantially less likely to be white and more likely to qualify for free and reduced-price lunch. These neighborhoods also tend to be more urban—denser in population—and have higher poverty rates and lower levels of educational attainment. One of the most striking differences is in rates of economic mobility: children from low-income families in CIS neighborhoods experience rates of upward mobility that are, on average, 0.35 standard deviations lower than those in non-CIS neighborhoods.

It is not simply that CIS schools are located in cities with higher poverty or greater diversity; rather, within a given commuting zone, CIS programs tend to serve the neighborhoods with the greatest need. These neighborhoods are more racially diverse, have lower incomes, and exhibit lower rates of economic mobility than other neighborhoods in the same labor market. In Figure II Panel *B*, we show that these differences persist even when comparing neighborhoods within the same commuting zone: the within-CZ estimates are very similar to the raw differences, indicating that CIS programs are systematically targeted toward the more disadvantaged communities in each area, consistent with the program's mission.

Although CIS programs are more common in high-poverty neighborhoods and overall coverage is substantial, most neighborhoods—including many high-poverty ones—do not have a CIS school. In Figure II Panel *C*, we plot the relationship between poverty rates and the coverage of CIS programs in those neighborhoods. Among the tracts with the lowest poverty

rates, only 10-15% of tracts are served by a CIS program. In tracts with the highest poverty rates, this proportion rises to around 30%. This underscores both the targeted nature of CIS and the substantial scope for expansion, something we return to in the Conclusion. In the next section, we describe the data we use to estimate the impact of CIS on student outcomes.

III Data

We leverage three main sources of data for our analysis. First, we use administrative education records from Texas that track student outcomes in both the short and long run. Second, we incorporate administrative data from Communities In Schools (CIS) in Texas, which provide detailed information on student-level interactions with CIS navigators. Third, we use Census and IRS records to estimate the long-run effects of CIS on earnings outcomes outside of Texas. In this section, we describe each of these data sources in more detail.

III.A Texas Education Research Center

Our primary data come from administrative academic, postsecondary, and labor market records maintained by the Texas Education Research Center (ERC) at the University of Texas at Austin. These data link individual-level records from the Texas Education Agency (TEA), the Texas Higher Education Coordinating Board (THECB), and the Texas Workforce Commission (TWC). Critically, the ability to track students across these datasets allows us to follow a single individual from elementary school through high school, into college, and into the labor market. A more detailed description of the data construction can be found in Appendix A.

We begin with student-level data from TEA on all students enrolled in Texas public schools between 1995 and 2019. The dataset includes more than 109 million student-year observations for approximately 16 million unique students. For each student, we observe time-invariant characteristics—such as birth cohort, race/ethnicity, and sex—as well as time-varying information including campus attended, grade level, FRPL status, and indicators for being classified as "at-risk" under TEA guidelines.[7]

We also identify whether each student ever attends a school served by the CIS program. To do this, we worked directly with Communities In Schools to obtain a list of all schools with a CIS program as of the 2019–2020 school year, including the year the program began at each school. Using this information, we construct an indicator for whether a student ever attends a CIS

[7]The at risk indicator flag students who are at risk of dropping out of school, according to TEA. There are 13 criteria, including poor academic performance, interactions with the criminal justice system, parenthood/pregnancy, and experiencing homelessness. See Texas Education Agency (2013).

school and track the timing of CIS entry at each site. Of the 16 million students in our dataset, approximately 3 million ever attend a school that has adopted CIS by 2019.

We construct a range of in-school outcomes for each student. Our primary cognitive outcome is performance on standardized state math assessments, administered annually in grades 3-8.[8] Because standardized tests are not administered universally in high school and our empirical strategy requires observing students prior to their exposure to CIS, we focus our analysis of test score outcomes on middle school. We also examine behavioral outcomes, including attendance and discipline. For attendance, we observe the number of days present in each of the six, six-week grading periods of the school year. For discipline, we construct indicators for whether a student was ever suspended (in-school or out-of-school) during the year, as well as a continuous measure of disciplinary incidents based on severity and frequency[9].

We follow students over time to measure long-run educational outcomes. For each student, we observe whether they graduate from high school by 2022. We also track in-state college enrollment using records from the Texas Higher Education Coordinating Board (THECB) through 2019.[10] We define indicators for whether a student ever attends any post-secondary institution between ages 18 and 22, and separately for two-year and four-year colleges.

We measure labor market outcomes using quarterly earnings data from the Texas Unemployment Insurance (UI) system, which capture wages for UI-covered employment but exclude self-employed and informal workers. We impute zero earnings in quarters without reported income and restrict the sample to individuals who ever appear in the UI data between ages 18 and 25, covering roughly 82% of students. We then construct earnings percentile ranks by age and cohort. However, because most of our Texas sample is too young to observe long-run earnings, we use these data primarily to validate results in the national sample and rely on linked Census and IRS records to measure earnings (see Section III.C).

Table I provides summary statistics from the ERC data for three groups: all students statewide, students who ever attend a school with a CIS program as of 2019, and students who attend a middle school with a CIS program in 2019. Our data include more than 16 million students across Texas, with approximately 3.3 million attending a school that ever implements CIS.

Students who ever attend a school that receives a CIS program by 2019 are on average somewhat more disadvantaged than the typical Texas student, but not dramatically so. They are

[8]The specific assessments vary over time (e.g., TAAS (1990-2002), TAKS (2003-2012), STAAR (2012-), so we standardize scores within grade and year. Eventually, the STAR exam switches from a grade based exam to a series of end of course exams, so when analyzing test scores we restrict to exams taken in 2016 or earlier to avoid this change.

[9]For each suspension incident we observe the type of suspension (in-school or out-of-school) as well as the number of days the student was required to serve

[10]The ERC also maintains National Student Clearinghouse records, which capture out-of-state enrollment, but only for more recent cohorts. We focus on in-state enrollment using THECB data to maximize coverage across the full panel.

less likely to be white (26.2% vs. 36.0%) and more likely to qualify for free or reduced-price lunch (52.2% vs. 47.0%). On average, they score about 0.13 standard deviations lower on standardized math tests. Long-run outcomes are broadly similar: high school graduation rates are nearly identical, and average earnings at age 27 differ by only $1,000 ($23,800 vs. $24,800 statewide). However, these average differences mask substantial heterogeneity in student need within schools, which we explore in Section IV. In particular, students who are case-managed are more likely to come from lower-income backgrounds and have markedly worse outcomes than their peers in the same school.

III.B CIS Student-Level Data

We use student-level data from Communities In Schools of Texas, available beginning in 2014, and link them to Texas ERC administrative records using student identifiers. These data allow us to observe individual students' participation in CIS and the specific services they receive. There are two key components: (1) intake records for students who are formally case-managed by CIS, which include baseline needs determinations, and (2) logs of each interaction between students and CIS navigators, detailing the date, duration, and type of service provided.

The participant data contain one row per student-year for students who were formally case-managed by CIS. For each student, we observe the date that parental consent was received, indicating permission for the student to participate in CIS programming and receive case management. We also observe the broad category of need that the students were flagged for. These include any combination of academics, attendance, behavior, and social services. In addition, we have more detailed information on the types of needs that navigators anticipate targeting. These include domains such as "academic readiness", "social skill", "self esteem", "classroom conduct", and "family conflict".

The interaction data contain one row per student-interaction with the navigator. For each interaction, we observe the date, the type of service provided, the organizational provider (whether the service was delivered by CIS staff or an external partner), and the duration of the interaction. For example, a navigator might spend 10 minutes providing food assistance through a backpack program, or one hour leading a workshop with a small group of students.

Combining these two data sources with the broader ERC administrative data, we can characterize both the types of students who receive case management from CIS and the nature of their interactions with navigators. Case-managed students are somewhat more likely to be female than male (46.6% male), are overwhelmingly from racial and ethnic minority groups (only 12.8% are white), and the vast majority qualify for free or reduced-price lunch (71.0%). Academically, these students tend to struggle: even at the 75th percentile, their standardized math scores fall

below the statewide mean, and the average score is more than half a standard deviation below the state average. Case-managed students also face higher disciplinary rates. In sixth grade, 26.5% receive an in-school suspension–more than twice the statewide average of 12.9% (see Appendix Table A.1 for details).

Consistent with these student characteristics, academics and behavior are the most common reasons for referral to CIS case management: 83.4% of students are referred for academic concerns and 65.5% for behavioral issues (not mutually exclusive). This focus is reflected in the specific needs that navigators aim to address. The most frequently targeted need is "test readiness," which appears in 41.0% of case-managed student-years. Behavioral supports are also prevalent: 36.0% of students are targeted for social skills, 32.3% for self-esteem, and 15.3% for issues related to classroom conduct. Appendix Table A.2 provides a full list of targeted needs.

These data also confirm that CIS provides services targeting both cognitive and non-cognitive domains. Analyzing the ten most common services provided in each year from 2015 to 2017, we find that time spent is roughly evenly split between cognitive and non-cognitive supports (see Appendix Figure A.1). The most common services include academic and social skills development, tutoring, character building, and attendance interventions. Notably, food assistance is also among the most frequently provided services, potentially influencing both academic performance and behavioral engagement.

III.C Census-IRS Data

In Section IV.C, we estimate the impact of CIS on long-run earnings outcomes using linked Census and IRS data. Our analysis relies on a combination of the 2000 and 2010 Census short forms and data from federal income tax returns in 1979, 1984, 1989, 1994, 1995, and 1998-2019. We follow Chetty et al. (2025*b*) to construct our analysis sample.[11] We focus on children born in the 1978–1992 birth cohorts who appear in the 2020 Numident file and are successfully linked to a parent via dependent claiming on 1040 tax forms.[12]

We measure both parent and child income using IRS tax records. Parent income is defined as the average of total household income between child ages 13–17, imputing zeros for any year without a tax filing. We exclude families with zero or negative mean income over the entire 5-year window. We rank parents relative to others with children in the same birth cohort to construct percentile ranks. For children, we measure household income at age 27—including spousal income, if present—and convert it to percentile ranks relative to all other children nationally in the same cohort.

[11] See Section II of Chetty et al. (2025*b*) for additional details on data sources and sample construction.

[12] Valid parents are adults who appear in the 2020 Numident and would have been between ages 15–50 at the time of the child's birth. The first such individual(s) is (are) assigned as the child's invariant parent(s).

Although we do not observe school enrollment in these data, we infer neighborhood of residence based on the Census tract from which a parent first files a 1040 tax return claiming the child. We assign each child to this invariant neighborhood and crosswalk Census tracts to school catchment zones using the procedure described in Section IV.C.

Our final sample includes 689,200 children who grew up in neighborhoods that ever received a CIS program during our analysis window. Of those, approximately 392,000 children would have been exposed to the program, based on their age at the time of implementation in their assigned school. We return to these data after detailing our empirical methodology and estimating the effect of CIS on short-run student outcomes in the Texas data.

IV Empirical Methodology

We turn now to developing a research design to estimate the effects of CIS on student outcomes in the short- and long-run. If CIS programs had been randomly assigned, we could identify causal effects by simply comparing outcomes in schools that did and did not receive the program. In practice, adoption was not randomized: local CIS affiliates select schools based on persistent need as well as funding, staffing, and other logistical constraints. We therefore use a difference-in-differences approach, comparing changes in outcomes within CIS schools following program entry to changes over time in a comparison group of similar schools that have not or have not yet adopted the program.

While CIS programs operate at the school level, the most intensive supports–such as case management–are provided to a subset of students, as described in Section II. To better understand the effects of CIS on the students most likely to be served by the program, we construct a proxy for case-managed status at the student level using pre-treatment student characteristics. While we have actual case-management status in later years, we do not use this label directly as this would risk conditioning on negative, contemporaneous shocks that both lead students to be case-managed and directly impact student outcomes.

As a starting point, we assess how well individual pre-treatment variables predict case management status using simple, univariate comparisons. For each student who attends a CIS middle school between 2014 and 2019, we construct an indicator for ever receiving case management during middle school, using the CIS microdata. In Figure III Panel *A*, we show that 19.1% of students in the bottom 5% of the elementary school standardized math score distribution are ever case managed in middle school, compared to 4.5% of students in the top 5% of the distribution. Similarly, students with the highest number of elementary school absences are nearly twice as likely to be case-managed, compared to students with near-perfect attendance. These univariate predictors show clear targeting patterns, but explain only a modest share of the variation

in case management status on their own.

To improve predictive accuracy, we train a gradient boosting model using the full set of elementary school predictors that we observe (Ke et al., 2017). Inputs include standardized math scores, attendance, disciplinary history, grade retention, and demographic variables such as race and ethnicity, sex, and FRPL status. The model is trained using five-fold cross-validation and evaluated on a held-out test set (see Appendix B for additional details). In Figure III Panel *C*, we show the out-of-sample accuracy of the model. For each percentile of the prediction distribution, we plot the fraction of students in the hold out data who were ever case-managed. The model has an AUC of 0.71, meaning if we were to randomly select a student who was case managed and a student who was not case managed, the model would assign a higher score to the true case managed student 71% of the time. By including the full set of elementary predictors, we achieve greater predictive power than any individual variable: among the 5% of students with the lowest predictions, only 1.1% are case managed, while in the top 5% of predictions, 37.4% are. Figure III, Panel *D*, shows the relative importance of each predictor based on model "gain," a measure of the improvement in model fit attributed to splits on that variable. Test scores are among the most important predictors, but other features—including attendance, grade retention, and prior disciplinary incidents—also contribute meaningfully.

Because our goal is to estimate the effect of CIS on students most likely to be served by the program, we apply the trained prediction model to the full panel of students—across all years and all schools—regardless of whether the school has implemented CIS. This allows us to construct a measure of predicted case management risk for every student in the data, even in years and schools where CIS was not active. Throughout the paper, we categorize students in the top quartile of the prediction distribution as "high-risk" and students in the bottom quartile as "low-risk" (as shown in Figure III Panel *C*). In treated schools, this enables us to identify high-risk students before the program begins. In control schools, it allows us to identify observationally similar high-risk students who were not exposed to CIS.

Using these predictions, we report summary statistics for high- and low-risk students in both CIS and non-CIS schools. Table II presents these results. We find large differences between high- and low-risk students across both groups of schools. For example, 74.9% (77.6%) of high-risk students are eligible for FRPL in a given year, compared to 9.2% (7.1%) of low-risk students in CIS (non-CIS) schools. Disparities in outcomes are similarly large: high-risk students score 1.4 standard deviations lower than low-risk students in 6th grade and earn half as much at age 27 ($16,000 vs. $34,000). Although outcomes for these groups are similar in CIS and non-CIS schools, the concentration of high-risk students is substantially higher in CIS schools. High-risk students outnumber low-risk students by a factor of 1.68 in CIS schools, compared to 0.91 in non-CIS schools.

Having constructed a measure of predicted treatment intensity for each student, we now turn to estimating the effect of CIS on student outcomes. We implement an event study framework that compares outcomes in treated schools in the year before and after CIS adoption to a set of control schools that never receive CIS. We measure outcomes at the student-year level (e.g. test scores in 6th grade) and define event time relative to the first year of program implementation. We estimate treatment effects in the eleven year window centered around the program beginning. Because we use elementary school characteristics to predict case management risk, we focus our analysis on programs beginning in middle schools.

We select a matched comparison school for each treated school based on pre-treatment characteristics. For each treated school, we measure covariates six years prior to the start of its CIS program (i.e., at event time $t = -6$), ensuring that the match is based on data outside the event study window. For potential control schools, we compute the same covariates at $t = -6$ relative to a hypothetical start year, aligning the match in calendar time. To operationalize the match, we estimate a linear probability model of CIS adoption as a function of school-level demographics and outcomes, using data from the donor pool of non-treated schools. We then compute the predicted probability of treatment for each school and match each treated school to the control school with the closest predicted value.[13] Because a single control school may serve as a match for multiple treated schools—either within the same implementation cohort or across implementation cohorts—we cluster all standard errors at the school-by-birth-cohort level in our baseline estimates.[14]

After matching, we stack the treated and control school pairs across implementation cohorts and estimate the following event study specification:

$$y_{i,t} = \sum_{\tau \in [-5,5]; \tau \neq -1} \beta_\tau \cdot (\mathbf{1}[t - T(s) = \tau] \cdot CIS_s) + \lambda_{s,T(s),r,m} + \gamma_{t,T(s)} + \varepsilon_{i,t}, \tag{1}$$

where $y_{i,t}$ is an outcome for student i in year t who attends school s with implementation cohort $T(s)$ (i.e. the year in which school s adopts CIS, or the pseudo-year for the never treated control group). The coefficients β_τ capture the difference in outcomes in treatment and control schools at relative time τ, with $\tau = -1$ as a reference period. The fixed effects $\lambda_{s,T(s),r,m}$ absorb level differences across schools by implementation cohort, student i's race r, and sex m. The $\gamma_{t,T(s)}$ fixed effects control for year-specific shocks within each implementation cohort, capturing the

[13] The full set of predictors includes the share of students eligible for FRPL; the share identified as at risk of dropping out (per TEA); the racial/ethnic composition (White and Hispanic shares); the fraction of students in the top and bottom quartiles of the predicted case management distribution; the share of students chronically absent (missing more than 10% of the school year); in-school suspension rates; standardized math scores; and the predicted high school graduation rate, described in Section VI.

[14] In robustness checks, we find that our results are similar when expanding the set of control schools to the nearest 5 matches or restricting to unique control matches. We also present results using more conservative school-level clustering in Appendix Table A.3.

evolution of outcomes in the control schools.

In many cases, we also report results from a more parsimonious difference-in-differences specification:

$$\begin{aligned} y_{i,t} = &\beta_{post,1-2}(\mathbf{1}[1 \leq \tau \leq 2] \cdot CIS_s) \\ &+ \beta_{post,3-5}(\mathbf{1}[3 \leq \tau \leq 5] \cdot CIS_s) \\ &+ \tilde{\lambda}_{s,T(s),r,m} + \tilde{\gamma}_{t,T(s)} + \varepsilon_{i,t}. \end{aligned} \tag{2}$$

This specification includes the same school and year fixed effects as Equation 1, but instead of estimating separate coefficients for each relative year, it averages the treatment effect over two post-treatment periods: years 1–2 and years 3–5. We focus primarily on $\beta_{post,3-5}$, which captures the average effect several years after implementation.

The key identifying assumption in this design is that, absent the introduction of CIS, outcomes in the treated schools would have followed the same path as those in the matched control schools. This assumption may be violated if the timing of CIS coincides with other reforms or resource shifts at the school (e.g., leadership changes, increases in funding) that also impact student outcomes. It could also be threatened if the composition of students changes at the time of CIS adoption, for example due to differential sorting.

In the next section, after presenting our baseline results, we assess the robustness of these findings to a range of alternative specifications designed to address these concerns. We begin by estimating the effect of CIS on student outcomes in the matched sample, then test sensitivity to alternative research designs—including using later-treated schools as controls, restricting comparisons within commuting zones, and comparing high- and low-risk students within CIS schools. We also examine whether other school characteristics, such as student demographics or teacher composition, change alongside CIS entry.

V Effect of CIS on Student Outcomes

We begin by documenting the first stage of our analysis: the impact of CIS adoption on students' exposure to the program. For each student-year, we calculate cumulative years of exposure to a CIS program during middle school. We estimate Equation 1 separately for high- and low-risk students, comparing changes in exposure in treated versus matched control schools. Figure IV presents the event study results. Before program adoption, there are no differences in exposure, as expected.[15] Following CIS entry, exposure rises rapidly and then levels off, consistent with

[15] Although we measure exposure during middle school, the CIS program does not begin in treated schools until $t = 0$, and control schools are never treated, exposure is not mechanically 0 for students in the pre-period as they could

students progressing through middle school grades in CIS schools. In the absence of student mobility across schools or grade retention, we would expect exposure to equal one year at $\tau = 0$, increase to approximately 1.67 at $\tau = 1$, and level off at 2.0 in subsequent years.[16] The observed first stage closely tracks this pattern. At $\tau = 1$, students average 1.45 (high-risk) and 1.56 (low-risk) years of exposure, roughly 90% of the anticipated exposure level.

In the years three to five after adoption—the main period we use for estimating more parsimonious DID effects throughout our analysis—students average 1.70 years of exposure (high-risk) and 1.74 years (low-risk). These estimates are statistically indistinguishable (t-statistic on difference = -1.37; see Table III), confirming that high- and low-risk students experience similar exposure to CIS at a school-level once the program is in place.

V.A Impacts on Test Scores

Having documented the relationship between CIS and total program exposure, we now turn to measuring the impact of that exposure on students' standardized test scores in middle schools. We start by estimating Equation 1 separately for high-risk students, with results shown in Figure V Panel *A*. The overall patterns mirrors the first stage. In the pre-period, coefficients are centered around zero, supporting the parallel trends assumption. Recall that control schools were matched on covariates measured at $\tau = -6$, outside the event-study window. Thus, the pre-period serves as an out-of-sample test of parallel trends between treated schools and the matched control schools.

Following CIS adoption, test scores begin to improve in treated schools relative to their matched controls, before stabilizing in later years. We estimate a difference-in-differences effect of 0.11 standard deviations (S.E. = 0.02) for high-risk students. For context, this magnitude is similar to the long-run impact of well-known education interventions. Prior work on the Tennessee STAR experiment has found that assignment to a small class in third grade led to gains of roughly 0.1 standard deviations on later standardized tests (Athey, Chetty and Imbens, 2025). When we scale the effect using an instrumental variables (IV) specification based on years of exposure, we find that three years of CIS exposure increases test scores by 0.18 standard deviations (S.E. = 0.04).

Off-the-shelf standard errors in our setting do not account for the fact that we estimate students' risk status using a predictive model, which introduces an additional source of uncertainty into our second-stage estimates. In Figure V Panel *A*, we find a t-statistic of 4.67, implying that standard errors would need to be approximately 140% larger in order to fail to reject a positive

have attended a different middle school with a CIS program prior to enrolling in either the treatment or control school. In practice, however, at $t = -1$, average exposure is less than 0.03 years. In Appendix Figure A.7, we also show there is essentially no difference in exposure to CIS in elementary school between these groups.

[16] If students never move and there are an equal number of students in 6-8th grade in a school, then in any given year at $\tau >= 3$, 6th graders will have one year of exposure, 7th graders 2, and 8th graders 3, averaging to 2 years of exposure.

treatment effect at the 5% level. Even with more conservative school-level clustering (as opposed to school-by-birth-cohort) the standard errors would need to be nearly 40% larger in order to fail to reject. To assess the extent to which this uncertainty might affect inference, we conduct a clustered bootstrap procedure in which we sample schools with replacement and, in each iteration, re-estimate both a simplified version of the risk model and the treatment effect. The distribution of the bootstrapped treatment effects is shown in Appendix Figure A.3. The resulting 95% confidence interval is 10.4% wider than that implied by the analytic standard errors clustered at the school level.[17] While this adjustment modestly widens the interval, the estimated impact of CIS on test scores for high-risk students remains statistically significant.

How do the meaningful effects of CIS on test scores for high-risk students compare to the effects for their lower-risk peers? In Figure V Panel *B*, we repeat the analysis for low-risk students. As with high-risk students, we observe coefficients close to zero in the pre-period, supporting the parallel trends assumption. However, following CIS adoption, we find no evidence of improvement in test scores for low-risk students; the estimated treatment effect is 0.01 standard deviations (S.E. = 0.02). In Table III, we report the DID coefficients from a fully interacted joint specification, estimating the treatment effects for high- and low-risk students simultaneously. We can reject that treatment effects for high- and low-risk students are the same (t-statistic = 3.23).

To assess whether the treatment effects decline gradually across the risk distribution—or are concentrated entirely among high-risk students—we repeat this analysis for the middle two quartiles of the predicted risk distribution. We find a monotonic pattern: treatment effects are largest for high-risk students, somewhat smaller for students in the third quartile, smaller still for those in the second quartile, and near-zero for low-risk students (see Appendix Figure A.2). This gradient suggests that CIS has the greatest impact on the students most likely to be served by the program.

While high-risk students are more likely to receive case management, we caution against interpreting the difference in treatment effects across groups as the causal impact of case management itself. One possibility is that low-risk students see no gains because they are unlikely to be directly served by the navigator. However, it may also be that CIS services—whether case management or broader supports—are less impactful for students with fewer needs. These estimates should therefore not be read as evidence that case management has no effect for the marginal low-risk recipient. For this reason, we do not rescale the reduced-form estimates by case management take-up, and instead emphasize the central finding: CIS meaningfully improves outcomes for high-risk students.

[17] For computational reasons, we do not re-estimate the full GBM model and instead use a somewhat simpler model, which we describe in the notes of Appendix Figure A.3. However, we keep this internally consistent for comparison between the bootstrap and analytic standard errors.

We perform several exercises to test the robustness of this result. First, we allow for a more flexible set of controls. We augment Equation 1 so that the year-specific shocks in γ can also flexibly vary by race/ethnicity and sex. This allows for the possibility that different demographic groups experienced distinct trends in test scores in the control schools and ensures tighter, "within-group" comparisons between treatment and control. The results are shown in Figure VI Panels *A* (high-risk) and *B* (low-risk). We again find clear, positive treatment effects for high-risk students and smaller, not statistically significant effects for low-risk students.

Next, we implement an alternative empirical strategy that compares earlier-treated schools to later-treated schools, rather than to matched control schools. For each implementation cohort $T(s)$ and relative time period τ, we define the set of control schools as those with later implementation cohorts ($T(s') > T(s) + \tau$).[18] We estimate a reduced form regression analogous to our baseline specification, using the same λ and γ fixed effects as in the baseline specification.

The results are shown in Figure VI, Panels *C* (high-risk) and *D* (low-risk). This approach yields similar treatment effects: for high-risk students, we estimate a 0.09 standard deviation increase in test scores (S.E. = 0.01), compared to 0.11 (0.02) in the baseline; for low-risk students, the estimate is 0.02 (0.02), similar to the baseline effect of 0.01 (0.02). Unlike our baseline design, which uses never-treated schools as controls, this approach exploits only the quasi-random timing of CIS adoption across eventually treated schools.

We also refine the later-treated design by restricting to comparisons within the same commuting zone. Specifically, we interact the γ year fixed effects with commuting zone indicators and compare earlier- and later-treated schools within the same city over time. Doing so ensures that both treatment and control schools are subject to the same commuting zone-level policies, funding environments, and initiatives. We continue to define the control group dynamically as schools with later implementation dates, as in the previous specification. Figure VI, Panels *E* and *F*, show that the results remain stable: high-risk students see meaningful improvements in test scores, while low-risk students show modest gains. In this specification we can reject a null effect for low-risk students (t-statistic = 2.4), but consistent with the previous specifications, the treatment effect for high-risk students is more than twice as large (0.078 vs. 0.036). This within-city design provides additional support for the claim that the effects we observe are driven by the CIS program rather than differential local trends that coincide with the timing of CIS entry.

A related concern is that changes in student or teacher composition—rather than the CIS program itself—drive the improvements in outcomes we observe. To assess this possibility, we test for changes in observable school-wide characteristics at the time of CIS adoption. Figure VII presents the results. In Panel *A*, we plot an event study of the fraction of students who are

[18] Because control schools eventually receive treatment, the panel is not balanced: control schools drop out of the comparison group in a given relative year once they begin treatment themselves.

White, finding no evidence of any changes in racial composition around program entry. In Panel *B*, we extend this analysis to a broader set of student and teacher characteristics, reporting difference-in-differences estimates in standard deviation units for comparability. Across nearly all characteristics—including student demographics, teacher experience, and staff size—the estimated effects are small and statistically indistinguishable from zero. The only exception is average teacher pay, which falls by 0.12 standard deviations (approximately $800). Finally, we show that test score gaps between high- and low-risk students converge within treated schools—but were widening in control schools during the same period—further ruling out unobserved school-wide changes as an alternative explanation (see Appendix Figure A.5).

Although we show that observable student characteristics do not change when CIS is introduced, a remaining concern is that higher-ability, high-risk students may selectively sort into CIS schools at program entry. To address this, we construct an alternative assignment measure that predicts middle school enrollment based on students' elementary school rather than their actual middle school.[19] We then re-estimate treatment effects using this predicted school assignment, holding treatment and control matches fixed. As shown in Figure VI Panels *G* and *H*, the results remain similar: we continue to find gains for high-risk students and no effects for low-risk students. Since the first stage is somewhat weaker in this exercise driven by moves during childhood, we also report IV estimates based on three years of exposure for both actual and anticipated school assignments (Appendix Figure A.4). The similarity of the estimates across approaches helps alleviate concerns about endogenous sorting into CIS schools.

Finally, we show that these patterns are not specific to math or driven by changes in test-taking behavior. Reading scores exhibit a similar pattern, with sizable, albeit smaller, gains for high-risk students and no meaningful change for low-risk students. After CIS enters a school, reading scores rise by 0.05 standard deviations for high-risk students (S.E. = 0.02) and remain essentially unchanged for low-risk students (–0.006, S.E. = 0.02). We also test whether CIS affects the likelihood that students take the standardized exams. For each student-year, we construct an indicator for missing a standardized math score and estimate treatment effects on this outcome. High-risk students in treated schools become slightly less likely to be missing test scores (–0.4 pp), but the change is not statistically significant (t = –0.9). We similarly find no detectable effect for low-risk students. These results indicate that the achievement gains we document are not driven by CIS changing the composition of test-takers in treated schools (see Appendix Figure A.6).

Taken together, these results provide consistent evidence that CIS improves academic outcomes for high-risk students. Across all specifications, we find positive effects for high-risk students and no or modest gains for their low-risk peers. These finding suggest that CIS delivers

[19] We construct a transition matrix from grade g to grade $g+1$ separately by birth cohort and assign students to schools based on modal transitions from their first observed elementary school.

targeted academic benefits to the students most likely to be served the program. Moreover, the narrowing of achievement gaps within schools following CIS adoption reinforces the interpretation that the program drives these improvements, rather than broader shifts in school composition or environment. In the next section, we examine whether these short-run gains in test scores translate into improvements in long-run educational outcomes.

V.B Impacts on Long-Run Outcomes in Texas

In addition to improving academic performance, CIS aims to increase high school graduation and economic opportunity, particularly for students at risk of dropping out. In this section, we examine whether the gains in test scores for high-risk students translate into improvements in students' long-run outcomes. We focus on high school graduation as a key milestone, followed by post-secondary enrollment, and labor market outcomes.

Unlike test scores and other contemporaneous outcomes, long-run outcomes are only observed once per student. As a result, our empirical design must shift from capturing year-specific exposure to capturing a student's cumulative exposure to the CIS program. Rather than defining a single implementation cohort for each school, $T(s)$, as in Section IV.A, we define a grade-specific cohort, $T(s,g)$, as the first year in which students in school s and grade g are ever exposed to CIS during any part of middle school.

As an example, consider a middle school that adopts a CIS program in 2005. The students who were in 8th grade in 2004 complete middle school before the program begins and are therefore never treated. In contrast, the 7th graders in 2004 are exposed to CIS during their 8th grade year, and the 6th graders in 2003 (and 2004) are exposed in 7th (and 8th) grade. In this case, we define $T(s,8) = 2005$, $T(s,7) = 2004$, and $T(s,6) = 2003$. This cohort-specific adjustment allows us to estimate the following grade-specific version of Equation 1 that accounts for cumulative exposure to CIS:

$$y_i = \sum_{\tau \in [-5,5]; \tau \neq -1} \beta_\tau \cdot (\mathbf{1}[t - T(s,g) = \tau] \cdot CIS_s) + \lambda_{s,T(s),r,m,g} + \gamma_{t,T(s),g} + \varepsilon_i. \tag{3}$$

This approach is analogous to estimating treatment effects using a single-grade sample (e.g., restricting to students observed in 8th grade), and stacking separate single-grade specifications into a single, pooled regression framework. Our results are unchanged if we restrict the sample to 8th graders only.

We begin by estimating a modified version of the first stage, measuring cumulative exposure to CIS in elementary, middle, and high school, as well as total exposure across all grades. As our identifying variation is driven by programs that begin in middle schools, it is reassuring that

differences in exposure are concentrated at that level. In later years ($\tau >= 3$), students in treatment schools receive only modest additional CIS exposure in elementary school—on average, 0.12 more years, or just 5% of the treatment–control difference in middle school exposure—largely due to nearby elementary schools adopting CIS in subsequent years. Importantly, there is no difference in high school exposure between treatment and control students. These patterns support the interpretation that our estimated effects are primarily driven by exposure during middle school (see Appendix Figure A.7). To account for the small differences in elementary exposure, when reporting IV estimates, we include exposure across all grades.

Next, we estimate the effect of CIS on high school graduation, separately for high- and low-risk students. Results for high-risk students are shown in Figure VIII Panel *A*. As with test scores, we find positive treatment effects: pre-period coefficients are near zero, and graduation rates rise following CIS adoption. We estimate a difference-in-differences effect of 3.1 percentage points (S.E.=1.0).[20] Rescaling for the effect of 3 additional years of total exposure, we find that CIS increases graduation rates by 3.4pp (see Table IV). This represents a 5.2% increase over the high-risk baseline graduation rate of 65%, and closes 13% of the pre-treatment gap in graduation rates between high- and low-risk students.

As with test scores, we find no corresponding increase in graduation rates for low-risk students. Figure VIII Panel *B*, shows slightly negative pre-trends that persist in the post-period. We estimate a small decline of 1.1 percentage points (S.E. = 0.7pp) off a high baseline graduation rate of 90.8%. While we cannot reject a null or small positive effect for low-risk students, we can reject impacts as large as those observed for high-risk students (see Table III).

Together, these results show that CIS improves high school graduation rates—delivering on one of the program's central goals. The effects are concentrated among the students most likely to be served by the program, mirroring the pattern observed for test scores. By providing high-risk students with targeted support during middle school, CIS meaningfully alters their educational trajectory, leading to lasting improvements even after students leave the program's direct reach.

We next test whether the gains in graduation rates translate into increased post-secondary enrollment in Texas. We focus on 2-year college attendance, since students on the margin of graduating high school are more likely to attend 2-year rather than 4-year institutions. Figure VIII Panel *C* shows the results for high-risk students. The pattern resembles that of test scores and graduation: coefficients are flat in the pre-period and rise modestly after CIS adoption, though the estimates are less precise. We estimate a 2.7 percentage point increase in the share of high-risk students who ever enroll in a 2-year college between ages 18 and 22, but the standard error is large

[20] We define graduation as ever appearing as a high school graduate, but are results are not sensitive to this definition. If anything we find slightly larger for graduating by age 18 (3.4pp, SE=1.2) and our results are unchanged if we restrict to graduation between age 16 and 20.

(1.8pp), and the effect is not statistically significant. In contrast, we find no increase for low-risk students (Panel *D*).

Consistent with 2-year college being the primary margin of response, we find little evidence of effects on other types of postsecondary enrollment. For high-risk students, the estimated effect on attending any institution in the Texas Higher Education Coordinating Board (THECB) data is nearly identical to the effect on 2-year enrollment (2.9pp vs. 2.7pp; see Appendix Figure A.8). 31.5% of high-risk students attend college at some point between ages 18 and 22 and among those, nearly all of those students attend at least one 2-year college (97%). Only 9% of high-risk students ever attend at least one 4-year college in the pre-period, and we find no impact of CIS on 4-year college attendance. Low-risk students attend 4-year colleges at much higher rates (45%) but also show no treatment effect (DID = –0.06pp).

CIS increases educational attainment for high-risk students, both by raising high school graduation rates and through suggestive increases in 2-year college enrollment. To assess whether these educational gains translate into improved economic outcomes, we next examine impacts on earnings in early adulthood. Estimating earnings effects in our current Texas-based design requires a long panel: we must observe students in elementary school to predict risk and follow them for multiple decades to measure earnings at a meaningful age. There are only three birth cohorts for which we can measure in-school data in first grade, which we use in the risk prediction, and also earnings at age 27.[21] To overcome this limitation, we turn to a second design based on CIS programs outside of Texas, using linked Census and IRS data to measure labor market outcomes.

V.C Impacts on Earnings in National Sample

We estimate the impact of CIS on earnings using linked Census and IRS administrative data covering nearly the entire U.S. population. This shift in data and setting requires several changes to our empirical approach. First, we focus on CIS programs outside of Texas, which allows us to test whether the program's impacts generalize beyond the Texas context. Second, we use a staggered-adoption design that compares cohorts within treated and control neighborhoods (rather than treated and control schools), exploiting differences in CIS exposure from the timing of enrollment relative to program adoption. Finally, because the national data do not contain detailed early-life measures, we do not distinguish between high- and low-risk students in this analysis, though we explore heterogeneity by parent income.

A key challenge in the Census-IRS data is that we do not observe which schools children attend. Instead, we rely on addresses to map between neighborhoods and school catchment zones. We assign children to neighborhoods based on the first Census tract in which they are claimed by

[21] For example, a student in first grade in 1995—the first year of our education data—turns 27 in 2016. We can measure data over the entire window for the first-graders in 1995,1996, and 1997.

their parents.[22] We then map each tract to the set of schools that contain the tract in their catchment zones following the approach in Chetty et al. (2025*a*).[23]

Unlike in Texas, we do restrict attention to middle schools. We are no longer constrained by the availability of universal standardized tests or by the need to use elementary data to assign high- and low-risk status. Instead, we use programs at all levels of school and map CIS exposure to birth cohorts. For each tract ever served by CIS, we identify the first birth cohort exposed to CIS, based on the timing of the program's introduction. For example, if a tract contains both an elementary school that adopted CIS in 2002 and a middle school that adopted in 2000, we define the first treated cohort as the 8th graders enrolled in the middle school in 2000—primarily students born in 1987. We denote this first exposed cohort in tract n as the tract's implementation cohort, $C(n)$, analogous to our approach in Section IV.

We implement a staggered adoption design, comparing outcomes for children in tracts exposed to CIS earlier versus later. For each implementation cohort $C(n)$, we define the control group as the set of tracts where CIS is introduced more than five birth cohorts later—that is, all n' such that $C(n') > C(n) + 5$.[24] For these later-treated control tracts, we assign a pseudo-implementation cohort $\tilde{C}(n') = C(n)$, matching the implementation cohort of the treated group they are compared to. We then stack the data across implementation cohorts following the approach of Sun and Abraham (2021), fully interacting all specifications with indicators for the implementation cohort. This structure allows later-treated tracts to serve as controls for multiple earlier implementation cohorts. We cluster standard errors at the high school catchment level.

We estimate:

$$y_i = \sum_{\tau \in [-5,5]; \tau \neq -1} \beta_\tau \cdot (\mathbf{1}[c - \tilde{C}(n) = \tau] \cdot CIS_n) + \phi_{n,r,p,\tilde{C}(n)} + \delta_{c,r,p,\tilde{C}(n)} + \varepsilon_i, \tag{4}$$

where y_i is an adult outcome for individual i, who grew up in Census tract n, was born in cohort c, belongs to race/ethnicity group r, and whose parents are in income quartile p. The term $\tilde{C}(n)$ denotes the implementation cohort for the treated tract and the pseudo-implementation cohort for the matched control tract.

This structure closely parallels our baseline specification in Equation 1, adapted to a cross-cohort design. The fixed effects ϕ absorb neighborhood-by-demographic variation within

[22] This is analogous to the approach we take in Appendix Figure A.4, assigning students to schools based on their first elementary school, and helps to limit concerns of endogenous sorting into schools and neighborhoods with CIS programs.

[23] We use the same crosswalk between school catchment zones and Census tracts as in Chetty et al. (2025*a*), provided by Peter Bergman and originally from Maponics (2017).

[24] This restriction ensures that the sample remains balanced across all relative time periods in the event study window. In Section V.A and Figure VI, we do not impose this restriction. However, applying it here does not drive the results: the unbalanced panel yields similar conclusions.

each implementation cohort, analogous to the school-by-demographic fixed effects in Texas. The δ terms control for cohort-by-demographic shocks within implementation cohorts, analogous to year fixed effects. The coefficients β_τ trace the path of estimated treatment effects from five cohorts before to five cohorts after the CIS program begins in treated tracts.

We begin by estimating the first stage in this setting using Equation 4 with the total number of years of exposure to CIS as the dependent variable. The resulting event study coefficients are shown in Figure IX, Panel *A*. The pattern is similar to the one shown in Figure IV: exposure begins to rise after the program is introduced and then levels off. To interpret the magnitudes, consider the coefficient at event time $\tau = 0$, corresponding to the first cohort exposed to CIS. If no students moved away from their initial Census tract, this coefficient would mechanically equal 1.[25] Instead, we estimate a coefficient of 0.51, reflecting migration during childhood. Importantly, the maximum potential years of exposure in this design is not capped at three years, as in our Texas middle school setting, but can be somewhat higher depending on the number and levels (elementary, middle, high school) of schools in the neighborhood that adopt CIS. In practice, we estimate that children born 3–5 years after the implementation cohort—i.e., cohorts with the longest potential exposure—experience an average of 2 additional years of CIS exposure relative to children in later-treated control tracts.

Does this exposure translate into improved outcomes later in life? Figure IX Panel *B* plots the effect of CIS on household income rank in the national distribution at age 27. The estimated coefficients in the pre-period are close to zero and statistically insignificant, suggesting that treatment and control tracts were on parallel trajectories before CIS was introduced. In the years following implementation, children in the treated tracts begin to earn more than their peers in later-treated control tracts. Estimating the difference-in-differences analog of Equation 4, we find that CIS improves adult income ranks by 1.14 ranks (S.E. = 0.37) for students born 3-5 years after the implementation cohort, corresponding to a 3.2% increase in income—about $850 annually—relative to a pre-period average of $26,270. Rescaling by the first stage, we estimate that 3 years of exposure to CIS programs increases earnings at age 27 by 1.54 percentile ranks (S.E. = 0.45), equivalent to a $1,140 increase (4.3%) in annual earnings.

Although the earnings results are estimated using CIS programs outside of Texas, we do not believe these impacts are unique to that setting. As described in Section B, our Texas panel is not long enough to observe earnings ranks at age 27 directly. However, we can predict later-life earnings using earlier observed outcomes. We estimate a machine learning model—similar to the one used to predict case management in Section IV—to forecast income rank at age 27 based

[25] We measure exposure assuming that students attend their local schools, though this is not always the case. Chetty et al. (2025*a*) use the same tract-to-school crosswalk and show in North Carolina that it reliably predicts actual school attendance. Deviations from local attendance would attenuate the first stage, making our estimates an upper bound for exposure and any IV estimates a lower bound for the treatment effect.

on earnings ranks and college attendance between ages 18 and 22 (Athey et al., 2025; Chetty, Deming and Friedman, 2025). See Appendix Figure A.10 for details. Using these predictions as the dependent variable, we then estimate Equation 3. The results closely mirror those from Figure IX (see Appendix Figure A.9 for a direct comparison), suggesting that the earnings gains from CIS are similar both inside and outside of Texas.

In the Texas sample, we found large differences in treatment effects for children who more versus less likely to be intensely served by CIS. Although we cannot construct the same predictions of high- and low-risk in the national sample, we examine heterogeneity by parent income as a proxy for student need. Specifically, we estimate difference-in-differences treatment effects separately for children whose parents were in the top quartile (lower need) and bottom quartile (higher need) of the national income distribution. The broadly results mirror the patterns observed in Texas: we find a 1.14-rank increase in adult income (S.E. = 0.51) for children from low-income families, and no detectable effect for students from high-income families (–0.02 ranks, S.E. = 0.80), though the estimates are imprecise (see Appendix Figure A.11).[26] Mirroring our earlier findings, these results suggest that CIS exposure during childhood continues to benefit children with greater needs well into adulthood.

The earnings gains for low-income children are driven primarily by a lower likelihood of left-tail outcomes—such as non-employment or very low earnings—and a higher likelihood of holding a modestly paid job, rather than by increased chances of reaching the top of the income distribution. This pattern is consistent with CIS's mission to help students stay in school and avoid adverse outcomes such as dropping out. To investigate this, we construct four binary outcome variables indicating whether each child is in a given adult income quartile at age 27. We estimate difference-in-differences treatment effects on the probability of falling into each quartile. As shown in Figure X, exposure to CIS reduces the likelihood of being in the bottom quartile by 2.3 percentage points (S.E. = 1.1pp) and increases the likelihood of being in the second quartile by 1.7 percentage points (S.E. = 1.0pp). We find no significant effects on the probability of being in the top half of the adult income distribution. The bottom quartile comprises children earning less than $15,000 annually, but 80% of these children are not working. We cannot reject that the entire effect on bottom-quartile incidence is driven by changes in employment, proxied by the presence of W-2 earnings.

These patterns are consistent with the education effects documented in the Texas sample. CIS increases high school graduation rates for students in need and modestly boosts 2-year college attendance, but does not shift students into four-year college pathways. The earnings effects we document of $1,140 are approximately 17% of the returns to a four-year degree (Zimmerman,

[26] Unlike in Texas, where treatment effects decreased monotonically across risk groups, we observe somewhat larger effects among children from middle-income families, but we have limited power to compare effects across subgroups.

2014), yet they remain substantial. Using the lifetime earnings projection method from Hendren and Sprung-Keyser (2020), we estimate that three years of CIS exposure during childhood increase expected lifetime earnings by $\$36,000$ in present day value, corresponding to roughly $\$7,100$ in present value additional federal tax revenue per student.

The long-run earnings gains we estimate are sizable, but such impacts take decades to materialize—and by the time they are measurable, it is often too late to guide program design or policy decisions. This raises a natural question: could we have forecast these long-run effects using outcomes observed earlier in life? In the next section, we return to the Texas data to examine the extent to which effects on short-run outcomes explain the long-run effects we document. We also examine the role of personalization—a core feature of the CIS model—and test whether the variation in services across students is consistent with efficient allocation.

VI Mechanisms: Forecasting Long-Run Impacts and Personalization

Explaining Long-Run Impacts with Contemporaneous Outcomes:

In the prior section, we showed that exposure to CIS improves outcomes across the course of a student's life. Test scores rise soon after the program begins, high school graduation rates increase, and those same students go on to earn more in adulthood—often a decade or more after their initial exposure. But how much of this long-run impact can be attributed to CIS's effect on short-run outcomes? In other words, to what extent do improvements in academic performance, attendance, and behavior mediate the gains we observe later in life?

It is common practice to forecast long-run impacts by scaling up short-run effects on test scores (e.g., Guryan et al., 2023; Hanushek, 2011; Kline and Walters, 2016; You, Fu and Chou, 2025). For example, one could take the estimated effect of CIS on test scores from Section IV.A and extrapolate the corresponding impact on adult earnings using established relationships (e.g., Chetty, Friedman and Rockoff, 2014). This approach implicitly assumes that test scores are a valid *statistical surrogate* for long-run outcomes—that is, that the full effect of CIS operates through its impact on test scores (Athey et al., 2025). However, the projections will not be accurate if the surrogacy assumption is violated.

On one hand, projections based on test scores could overstate long-run effects. There are well-documented cases of "fade-out," where early gains in academic outcomes diminish over time (see Cascio and Staiger, 2012, for a review). For example, improvements in test scores might reflect increased teaching to the test or changes in classroom practices that boost scores without meaningfully improving long-term skills. Even if there are meaningful gains in human capital, they may be offset by later constraints—such as limited access to college counseling or labor market networks—that CIS does not fully address. In these cases, rescaling based on short-run test score

impacts would predict larger long-run effects than are ultimately realized.

On the other hand, projections based solely on test scores may understate the true long-run effects of CIS. This pattern has been documented in the context of several educational interventions, where modest impacts on test scores masked much larger improvements in adult outcomes (e.g., Anders, Barr and Smith, 2023; Chetty et al., 2011; Cortes, Goodman and Nomi, 2015; Gray-Lobe, Pathak and Walters, 2023). In such cases, test scores do not fully capture the dimensions of student development affected by the intervention. For example, long-run gains could additionally reflect changes in non-cognitive skills, engagement, motivation, or access to support systems that are not reflected in improved test scores. If CIS improves outcomes through these or other channels, then test scores alone may provide an incomplete picture of the true long-term impact.

Can we accurately predict the long-run effect of CIS using short-run outcomes? We study this question in the context of high school graduation because, in our Texas sample, we observe many cohorts with both early academic measures and later graduation outcomes. To do so, we construct predicted graduation outcomes using students' contemporaneous academic and behavioral indicators, and then compare the treatment effect of CIS on these predicted outcomes to the actual long-run effect.

We train a gradient boosting algorithm separately in grades 6–8, using data on all students, not just those in schools with CIS programs. The model includes a set of year-specific covariates: student demographics, standardized math scores, attendance, and disciplinary outcomes (including in- and out-of-school suspensions and the number of infractions). These predictors capture both academic outcomes (e.g., test scores) and behavioral outcomes (e.g., attendance and discipline), which may reflect non-cognitive traits relevant for long-run success. All predictors are measured contemporaneously, so the model output can be viewed as a composite measure of academic and behavioral performance in each year, with weights determined by how strongly each variable predicts high school graduation in the broader student population. See Appendix Figure A.12 for additional details.

We then estimate the impact of CIS on predicted high school graduation rates for high-risk students using our standard event study framework. The results are shown in Figure XI Panel *A*, alongside the actual effect on high school graduation for comparison. The patterns are remarkably similar: the increase in predicted graduation rates following CIS implementation closely mirrors the observed increase in true graduation rates.

In Figure XI Panel *B*, we present the difference-in-differences for these two outcomes. We estimate a true effect of CIS on high school graduation of 3.1 percentage points (S.E. = 1.0 pp), compared to a predicted effect of 2.4 percentage points (S.E. = 0.6 pp). In other words, nearly 80% of the true impact on high school graduation can be explained by the effect on observable,

contemporaneous outcomes.

Had we only used test scores, we would have substantially understated the effect of CIS on high school graduation. The results of two procedures are shown on the right hand side of Figure XI Panel *B*. In the first, we use the observational relationship between test scores and high school graduation rates. We fit a regression of graduation rates on test scores in each grade using statewide data and use that to predict graduation. In the second, we allow for a more flexible, possibly non-linear relationship by training a simplified gradient boosting model that uses only test scores as inputs. In both cases, the predicted impacts are substantially smaller: the predicted impacts from the GBM model are only 45% as large as the true effect.

More broadly, our findings highlight the potential of using rich, short-run administrative data to forecast long-run program impacts in real time. Test scores—and the cognitive skills they reflect—explain a meaningful share of CIS's long-run effects, but alone they understate its full impact. Behavioral measures like attendance and discipline also play a critical role. By combining these outcomes using only existing administrative records, we are able to predict long-run effects with a high degree of accuracy. This approach shows promise for evaluating interventions without waiting years or decades for follow-up data (Prentice, 1989; Athey et al., 2025).

Personalization in the CIS Model:

A defining feature of the program is that site coordinators tailor support to each student's specific needs—deciding, for instance, whether to prioritize academic tutoring, mentoring, counseling, or attendance interventions. This individualized approach could, in principle, be efficient or inefficient. If coordinators allocate services efficiently, students with different underlying challenges receive different mixes of support, but those bundles are well matched to their needs and yield similar overall improvements in long-run outcomes. If personalization is inefficient, however, some services may be more effective than others—so students who happen to receive the most impactful supports (for example, if tutoring were especially beneficial) would experience larger gains, while others assigned less effective services would benefit less. We explore whether the observed variation in service portfolios and outcomes is consistent with efficient matching of services to students' underlying needs.[27]

We begin by documenting that site coordinators tailor the support students receive to their individual needs. Figure XII Panel *A* plots the relationship between students' prior academic performance and the likelihood of being targeted for academic or behavioral support, controlling

[27]Formally testing efficiency would require quasi-experimental variation in service bundles. Our interpretation is therefore suggestive and rests on the assumption that, absent effective matching, average treatment effects would differ across services. Under this benchmark, the fact that students with different needs—who receive systematically different supports—experience similar long-run gains is consistent with coordinators allocating services to equalize overall impacts.

for the predicted probability of case management.[28] Students with lower elementary-school math scores are substantially more likely to be targeted for academic support: 90 percent of students in the bottom ventile of prior math scores are flagged for academics compared to 54 percent in the top ventile. Conversely, students with relatively strong academic performance—who appear in case management for other reasons—are more likely to be targeted for behavioral support. Students in the top ventile of math scores are more than 10 percentage points more likely to be flagged for behavioral issues than those in the bottom ventile.

These distinctions translate into meaningful differences in the set of services students ultimately receive. In Figure XII Panel *B*, we plot the share of services these students receive that are academic in nature.[29] Students with lower average standardized math scores in grades 3–5 receive a higher fraction of academic services than peers with stronger prior performance, even after conditioning on predicted case-management risk. This pattern holds within schools and years, indicating that coordinators actively customize services across students in a given cohort. Guided by this relationship, we classify high-risk students with below-median elementary-school academic performance as those with primarily academic needs, and high-risk students with above-median performance as those with primarily non-academic (e.g., behavioral or engagement-related) needs. We then compare the short- and long-run impacts of CIS across these two groups.

CIS has a larger impact on standardized test scores for students with primarily academic needs. We estimate a modified version of the difference-in-differences specification in Equation 2 for the high-risk sample, including a linear control for predicted case-management risk, an indicator for below-median elementary-school academic performance, and interactions of these controls with the post-period indicators. The resulting coefficients for each group are shown in Figure XII Panel *C*. The estimated effect on math scores for students with academic needs is nearly twice as large as that for students with primarily non-academic needs (0.103 vs. 0.056 SD; t-statistic on difference = 3.7).

Despite these differences in short-run academic outcomes, the long-run effects of CIS are nearly identical across the two groups. As shown in Figure XII Panel *D*, the estimated impacts on high school graduation are 0.028 and 0.029 percentage points, respectively. This pattern does not constitute definitive evidence of efficient personalization—unobserved heterogeneity could also explain the similarity in long-run effects—but we find no clear indication of inefficiency.

[28] Site coordinators assign students in the case-management database any combination of flags for academics, attendance, behavior, and social services. We define a student as flagged for academics (behavior) if their targeted areas include academics (behavior).

[29] The set of service codes in the case-management microdata varies slightly by year. As an example, in 2018–2019 we classify as academic services: Academic Readiness Interventions, Parent Conferences with Teachers, Academic Skills Development and Support, Credit Recovery Assistance, Academic Monitoring, Homework Assistance, Language (ESL/ELL) Support, Tutoring, Student/Teacher Conferences, Support for Students in Special Education, School Supplies, and General – Academic Enhancement and Support.

Students with different needs receive distinct service portfolios and experience different short-run impacts, yet achieve similar long-run gains. Still, it remains possible that students with primarily non-academic needs would have benefited even more if they had received a mix of services closer to those provided to students with academic needs. Ultimately, the data indicate that CIS's personalization yields distinct service portfolios across students without producing clear disparities in long-run efficacy.

VII Conclusion

Students from low-income families face persistent disadvantages in school and beyond (Chetty et al., 2014; Solon, 1992; Black, Devereux and Salvanes, 2005; Chetty et al., 2020). We show that providing personalized, coordinated supports in high-poverty schools can meaningfully improve outcomes– both in the classroom and over the long run. Communities In Schools (CIS), the largest integrated student support program in the country, raises test scores, boosts high school graduation, and improves early-career earnings for students most at risk of falling behind. These impacts are not driven by a single academic channel, but by improvements in both cognitive and non-cognitive outcomes.

Our results suggest that social connection through a dedicated navigator can significantly improve children's outcomes, echoing findings from programs targeting adults in other domains (Bergman et al., 2024; Caspi and Rafkin, 2025; Espinosa et al., 2024; Finkelstein and Notowidigdo, 2019; Katz et al., 2022; Weiss et al., 2019). In schools, this relational model not only improves students' outcomes, but does so in a relatively cost effective way. We estimate that each $1,000 invested in CIS raises adult earnings by approximately $406, compared to an estimated $40 per $1,000 from class-size reductions (Chetty et al., 2011).[30]

Although CIS now serves over 2 million students nationwide, our findings suggest there is still considerable scope for expansion—even in a state like Texas, where coverage is already high. We find that the benefits of CIS are concentrated among the most disadvantaged students, while effects are minimal for their more advantaged peers. This pattern provides a natural framework for guiding future expansions. In Figure XIII, we simulate a targeted rollout strategy that prioritizes schools with the largest concentrations of high-risk students. While CIS programs currently reach about 20% of middle school students in Texas, we estimate that the current allocation captures only 60% of the potential impact achievable at that coverage level. Expanding to 40% coverage—focusing on

[30]Chetty et al. (2011) estimate that 2.14 years of smaller class sizes raise earnings in adulthood by $368 and cost $9,355 per student. We estimate that 3 years of CIS raises earnings by $1,140. Although we do not observe costs directly, we cite Texas Legislative Budget Board (2016), which estimates conservatively that CIS costs $936 per case-managed student. Assuming that is equal to the cost per student per year, we estimate that $1,000 spent on CIS increases adult earnings by $406.

the next-highest-need schools—could achieve 92% of the maximum attainable impact on high-risk students' graduation rates. Our findings point to the potential for substantial additional gains from targeting intensive supports and personalized assistance to the students who need them most.

References

Aizer, Anna, Shari Eli, Joseph Ferrie, and Adriana Lleras-Muney. 2016. "The long-run impact of cash transfers to poor families." *American Economic Review*, 106(4): 935–971.

Alan, Sule, and Elif Kubilay. 2025. "Empowering adolescents to transform schools: Lessons from a behavioral targeting." *American Economic Review*, 115(2): 365–407.

Alan, Sule, Teodora Boneva, and Seda Ertac. 2019. "Ever failed, try again, succeed better: Results from a randomized educational intervention on grit." *The Quarterly Journal of Economics*, 134(3): 1121–1162.

Anders, John, Andrew C Barr, and Alexander A Smith. 2023. "The effect of early childhood education on adult criminality: Evidence from the 1960s through 1990s." *American Economic Journal: Economic Policy*, 15(1): 37–69.

Angrist, Joshua D, Victor Lavy, Jetson Leder-Luis, and Adi Shany. 2019. "Maimonides rule redux." *American Economic Review: Insights*, 1(3): 309–324.

Athey, Susan, Raj Chetty, and Guido Imbens. 2025. "The Experimental Selection Correction Estimator: Using Experiments to Remove Biases in Observational Estimates." National Bureau of Economic Research Working Paper 33817.

Athey, Susan, Raj Chetty, Guido W Imbens, and Hyunseung Kang. 2025. "The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely." *Review of Economic Studies*, rdaf087.

Attea, Kate, and Alan Tuck. 2005. "Communities In Schools: Propelling a National Network to the Next Level." The Bridgespan Group. Accessed December 3, 2025.

Barr, Andrew, and Benjamin Castleman. 2025. "Increasing Degree Attainment among Low-Income Students: The Role of Intensive Advising and College Quality." *American Economic Review*, 115(11): 4075–4103.

Bell, Alex. 2025. "The Long-Term Impacts of Mentors: Evidence from Experimental and Administrative Data." Working Paper, Georgia State University Department of Economics. Latest Draft: April 26, 2025.

Bergman, Peter, Raj Chetty, Stefanie DeLuca, Nathaniel Hendren, Lawrence F Katz, and Christopher Palmer. 2024. "Creating moves to opportunity: Experimental evidence on barriers to neighborhood choice." *American Economic Review*, 114(5): 1281–1337.

Bhuller, Manudeep, Magne Mogstad, and Kjell G Salvanes. 2017. "Life-cycle earnings, education premiums, and internal rates of return." *Journal of Labor Economics*, 35(4): 993–1030.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes. 2005. "Why the apple doesn't fall far: Understanding intergenerational transmission of human capital." *American Economic Review*, 95(1): 437–449.

Card, David. 1999. "The causal effect of education on earnings." *Handbook of Labor Economics*, 3: 1801–1863.

Cascio, Elizabeth U, and Douglas O Staiger. 2012. "Knowledge, tests, and fadeout in educational interventions." National Bureau of Economic Research Working Paper 18038.

Caspi, Aviv, and Charlie Rafkin. 2025. "Legal assistance for evictions."

Cattan, Sarah, Kjell G Salvanes, and Emma Tominey. 2025. "First-Generation Elite: The Role of School Social Networks." *American Economic Review*, 115(12): 4369–4403.

Chetty, Raj, David J Deming, and John N Friedman. 2025. "Diversifying society's leaders? The determinants and causal effects of admission to highly selective private colleges." *The Quarterly Journal of Economics*, qjaf050.

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014. "Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood." *American Economic Review*, 104(9): 2633–2679.

Chetty, Raj, John N Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2020. "Income segregation and intergenerational mobility across colleges in the United States." *The Quarterly Journal of Economics*, 135(3): 1567–1633.

Chetty, Raj, John N Friedman, Nathaniel Hendren, Maggie R Jones, and Sonya R Porter. 2025*a*. "The opportunity atlas: Mapping the childhood roots of social mobility." *The Quarterly Journal of Economics*, forthcoming.

Chetty, Raj, John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How does your kindergarten classroom affect your earnings? Evidence from Project STAR." *The Quarterly Journal of Economics*, 126(4): 1593–1660.

Chetty, Raj, Matthew O Jackson, Theresa Kuchler, Johannes Stroebel, Nathaniel Hendren, Robert B Fluegge, Sara Gong, Federico Gonzalez, Armelle Grondin, Matthew Jacob, et al. 2022. "Social Capital I: Measurement and Associations with Economic Mobility." *Nature*, 608(7921): 108–121.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. 2014. "Where is the land of opportunity? The geography of intergenerational mobility in the United States." *The Quarterly Journal of Economics*, 129(4): 1553–1623.

Chetty, Raj, Will Dobbie, Benjamin Goldman, Sonya R Porter, and Crystal Yang. 2025*b*. "Growing Class Gaps, Shrinking Race Gaps: Economic and Sociological Mechanisms Underlying Recent Trends in Intergenerational Mobility." *Quarterly Journal of Economics*, forthcoming.

CIS. 2011. "Communities In Schools: The Nation's Leading Dropout Prevention Organization." Communities In Schools, Inc. Technical Report, Arlington, VA. Retrieved from https://www.communitiesinschools.org/media/uploads/attachments/Communities_In_Schools_The_Nations_Leading_Dropout_Prevention_Organization_1.pdf.

CIS. 2023. "Communities in Schools: Databook 2021-22." *https://www.communitiesinschools.org/articles/article/2021-2022-communities-schools-data-book/*.

CIS. 2025*a*. "Our Approach." *https://www.communitiesinschools.org/our-approach/*, Accessed December 3, 2025.

CIS. 2025*b*. "Partner with a Local CIS Affiliate." *https://www.communitiesinschools.org/affiliates/*, Accessed December 3, 2025.

CIS. 2025*c*. "Site Coordinators." *https://www.communitiesinschools.org/site-coordinators/*, Accessed December 3, 2025.

CIS. 2025*d*. "Start a Local Affiliate." *https://www.communitiesinschools.org/network/starting-affiliate/*, Accessed December 3, 2025.

CIS of Houston. 2024. "Annual Financial and Compliance Report for the Year Ended August 31, 2023." *https://projects.propublica.org/nonprofits/download-audit?download=true&filename=2023-08-GSAFAC-0000030536*.

Corrin, William, Leigh Parise, Oscar Cerna, Zeest Haider, and Marie-Andrée Somers. 2015. "Case management for students at risk of dropping out: Implementation and interim impact findings from the communities in schools evaluation." *Available at SSRN 2609366*.

Cortes, Kalena E, Joshua S Goodman, and Takako Nomi. 2015. "Intensive math instruction and educational attainment: Long-run impacts of double-dose algebra." *Journal of Human Resources*, 50(1): 108–158.

Dobbie, Will, and Roland G Fryer Jr. 2011. "Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone." *American Economic Journal: Applied Economics*, 3(3): 158–187.

Doleac, Jennifer L. 2019. "Wrap-around services don't improve prisoner reentry outcomes." *Journal of Policy Analysis and Management*, 38(2): 508–514.

Espinosa, Javier, William N Evans, David C Phillips, and Tim Spilde. 2024. "How Do Holistic Wrap-Around Anti-Poverty Programs Affect Employment and Individualized Outcomes?" National Bureau of Economic Research Working Paper 32911.

Finkelstein, Amy, and Matthew J Notowidigdo. 2019. "Take-up and targeting: Experimental evidence from SNAP." *The Quarterly Journal of Economics*, 134(3): 1505–1556.

Fischer, Claude S, and Greggor Mattson. 2009. "Is America Fragmenting?" *Annual Review of Sociology*, 35: 435–455.

Goldman, Benjamin, Jamie Gracie, and Sonya R Porter. 2023. "Can Individualized Student Supports Improve Economic Outcomes for Children in High Poverty Schools?"

Gray-Lobe, Guthrie, Parag A Pathak, and Christopher R Walters. 2023. "The long-term effects of universal preschool in Boston." *The Quarterly Journal of Economics*, 138(1): 363–411.

Guryan, Jonathan, Jens Ludwig, Monica P Bhatt, Philip J Cook, Jonathan MV Davis, Kenneth Dodge, George Farkas, Roland G Fryer Jr, Susan Mayer, Harold Pollack, et al. 2023. "Not too late: Improving academic outcomes among adolescents." *American Economic Review*, 113(3): 738–765.

Hanushek, Eric A. 2011. "The economic value of higher teacher quality." *Economics of Education Review*, 30(3): 466–479.

He, Haibo, and Edwardo A. Garcia. 2009. "Learning from Imbalanced Data." *IEEE Trans. on Knowl. and Data Eng.*, 21(9): 1263–1284.

Hendren, Nathaniel, and Ben Sprung-Keyser. 2020. "A Unified Welfare Analysis of Government Policies." *The Quarterly Journal of Economics*, 135(3): 1209–1318.

Hyman, Joshua. 2017. "Does money matter in the long run? Effects of school spending on educational attainment." *American Economic Journal: Economic Policy*, 9(4): 256–280.

Jackson, C Kirabo, Rucker C Johnson, and Claudia Persico. 2016. "The effects of school spending on educational and economic outcomes: Evidence from school finance reforms." *The Quarterly Journal of Economics*, 131(1): 157–218.

Katz, Lawrence F., Jonathan Roth, Richard Hendra, and Kelsey Schaberg. 2022. "Why Do Sectoral Employment Programs Work? Lessons from WorkAdvance." *Journal of Labor Economics*, 40(S1): S249–S291.

Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, 3146–3154.

Kline, Patrick, and Christopher R. Walters. 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." *The Quarterly Journal of Economics*, 131(4): 1795–1848.

Lavecchia, Adam M, Philip Oreopoulos, and Robert S Brown. 2020. "Long-run effects from comprehensive student support: Evidence from pathways to education." *American Economic Review: Insights*, 2(2): 209–224.

Malkus, Nat. 2024. "Long COVID for Public Schools: Chronic Absenteeism Before and After the Pandemic."

Maponics. 2017. "Maponics School Boundaries."

McShane, Michael Q. 2019. "Supporting Students Outside the Classroom: Can Wraparound Services Improve Academic Performance?" *Education Next*, 19(3): 38–45.

MDRC. 2025. "Communities In Schools." *https://www.mdrc.org/work/projects/communities-schools*, Accessed December 3, 2025.

Nickow, Andre, Philip Oreopoulos, and Vincent Quan. 2020. "The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence." National Bureau of Economic Research Working Paper 27476.

Office of Head Start. 2025. "Head Start Program Facts: Fiscal Year 2024." *https://headstart.gov/program-data/article/head-start-program-facts-fiscal-year-2024*, Accessed: 2025-11-18.

Parise, Leigh, and Farhana Hossain. 2017. "Using Research Evidence to Strengthen Support for At-Risk Students: A Case Study of Communities in Schools (Issue Focus Infographic)." MDRC, New York, NY.

Parise, Leigh, William Corrin, Kelly Granito, Zeest Haider, Marie-Andrée Somers, and Oscar Cerna. 2017. "Two years of case management: Final findings from the Communities in Schools random assignment evaluation." *MDRC, Project: Communities In Schools*.

Porowski, Allan, and Aikaterini Passa. 2011. "The effect of communities in schools on high school dropout and graduation rates: Results from a multiyear, school-level quasi-experimental study." *Journal of Education for Students Placed at Risk*, 16(1): 24–37.

Prentice, Ross L. 1989. "Surrogate endpoints in clinical trials: definition and operational criteria." *Statistics in Medicine*, 8(4): 431–440.

Putnam, Robert D. 2016. *Our Kids: The American Dream in Crisis.* Simon and Schuster.

Robinson, Carly D, Biraj Bisht, and Susanna Loeb. 2025. "The inequity of opt-in educational resources and an intervention to increase equitable access." *Educational Researcher*, 54(6): 328–338.

Solon, Gary. 1992. "Intergenerational income mobility in the United States." *The American Economic Review*, 393–408.

Somers, Marie-Andrée, and Zeest Haider. 2017. "Using Integrated Student Supports to Keep Kids in School: A Quasi-Experimental Evaluation of Communities in Schools." *MDRC*.

Stevens, David W. 2007. "Employment that is not covered by state unemployment insurance laws." Center for Economic Studies, US Census Bureau.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*, 225(2): 175–199.

Texas Education Agency. 2013. "PEIMS Data Standards: At-Risk Indicator Code (E0919)." *http://ritter.tea.state.tx.us/peims/standards/1314/e0919.html*, Accessed: 2025-09-17.

Texas Education Agency. 2021. "Commissioner's Rules Concerning the Communities In Schools Program, Chapter 89, Subchapter EE." *https://tea.texas.gov/about-tea/laws-and-rules/commissioner-rules-tac/coe-tac-currently-in-effect/ch089ee.pdf*, Texas Administrative Code, Title 19, Part 2. Adopted July 13, 2021.

Texas Education Agency. 2022. "Communities In Schools." *https://tea.texas.gov/texas-schools/support-for-at-risk-schools-and-students/communities-in-schools*, Accessed: 2025-01-29.

Texas Legislative Budget Board. 2016. "Overview of Performance and Effectiveness Information for Texas Education Agency Non-Foundation School Programs." Texas Legislative Budget Board Presentation ID 3330, Austin, TX. Presented to the House Appropriations Committee, Subcommittee on Article III. See slide 12: "Average Expenditure Per CIS Participant" (state and local costs per case-managed student).

U.S. Department of Education, National Center for Education Statistics. 2022. "Table 309.30. Residence and migration of all first-time degree/certificate-seeking undergraduates in 4-year degree-granting postsecondary institutions who graduated from high school in the previous 12 months, by state or jurisdiction: Fall 2020." U.S. Department of Education. Digest of Education Statistics 2021, January 2022. Data source: Integrated Postsecondary Education Data System (IPEDS), Spring 2021, Fall Enrollment component. Available at https://nces.ed.gov/programs/digest/d21/tables/dt21_309.30.asp.

Weiss, Michael J, Alyssa Ratledge, Colleen Sommo, and Himani Gupta. 2019. "Supporting community college students from start to degree completion: Long-term evidence from a randomized trial of CUNY's ASAP." *American Economic Journal: Applied Economics*, 11(3): 253–297.

You, Shijun, Wei Fu, and Shin-Yi Chou. 2025. "Local Problems, Local Solutions: Evidence from the Drug-Free Communities Program on Youth Development." National Bureau of Economic Research Working Paper 34182.

Zimmerman, Seth D. 2014. "The Returns to College Admission for Academically Marginal Students." *Journal of Labor Economics*, 32(4): 711–754. Publisher: The University of Chicago Press.

TABLE I: Summary Statistics

	All Students (1)	Ever in School with CIS Program (2)	Ever in MS with CIS Program (3)
A. Student Characterisitcs			
Fraction Male	0.520	0.514	0.495
Fraction White	0.360	0.262	0.239
Fraction Hispanic	0.454	0.551	0.591
Fraction Black	0.142	0.157	0.143
Fraction Other Race/Ethnicity	0.044	0.030	0.028
Fraction Student-Years FRPL	0.470	0.522	0.515
Fraction Student-Years Dropout-Risk	0.450	0.497	0.511
B. Student Outcomes			
Standardized Math Score in 6th Grade	0.004	-0.133	-0.156
Absence Rate in 6th Grade	0.037	0.039	0.040
Fraction Ever Receive In-School Suspension in 6th Grade	0.129	0.169	0.169
Ever Ever Receive Out-of-School Suspension in 6th Grade	0.061	0.087	0.090
Fraction Ever Graduate from HS	0.738	0.732	0.727
Fraction Ever Enroll in TX College Ages 18-22	0.501	0.477	0.479
Fraction Appear in Earnings Data	0.820	0.846	0.833
Average Earnings at Age 27 ($)	24812	23818	23959
Employment Rate at Age 27	0.605	0.624	0.609
Average Individual Earnings Rank at Age 27	0.497	0.490	0.488
C. School Characteristics			
Average Number of Students	991	1368	903
Average Number of Teachers	65	91	61
Fraction Teachers Male	0.229	0.309	0.312
Fraction Teachers White	0.670	0.610	0.582
Fraction Teachers Hispanic	0.217	0.294	0.306
Fraction Teachers with Advanced Degree	0.266	0.280	0.258
Average Teacher Tenure (yrs)	7.643	7.584	7.428
Average Teacher Pay ($)	51586	52761	52027
Number of Student-Years (1,000s)	109172	27222	15547
Number of Students (1,000s)	16165	3332	1639
Number of Students in HS Grad Sample (1,000s)	9714	2248	1110
Number of Students in College Sample (1,000s)	8311	1907	911
Number of Students in Earnings Sample (1,000s)	5960	1351	614
Number of Schools	10682	613	208
Number of School-Years	189748	12941	4365

Notes: This table reports summary statistics for three groups: (1) all students in our data, (2) students who ever attend a school with a CIS program in 2019, and (3) students who ever attend a middle school with a CIS program in 2019. All statistics use the maximum available data spanning 1994 to 2019. Panel A summarizes student demographic characteristics. Panel B reports student outcomes. High school graduation is measured among students who are age 19 or older by 2019; the corresponding sample size and those for subsequent outcomes are listed at the bottom of the table. College enrollment is measured among students age 22 or older by 2019. The fraction of students in the earnings data refers to the share with non-zero earnings at any point between ages 18 and 25. The employment rate is the fraction with non-zero earnings at age 27, conditional on being in the earnings data. All earnings outcomes are measured among students who are age 27 or older by 2019. Panel C presents summary statistics on school characteristics. "Teachers with advanced degrees" refers to those with a master's degree or higher. "Teacher tenure" reflects years of continuous employment at the current school. All monetary values are expressed in 2015 dollars.

TABLE II: Characteristics of Low- and High-Risk Students

	Low-Risk in CIS School (1)	High-Risk in CIS School (2)	Low-Risk in Non-CIS School (3)	High-Risk in Non-CIS School (4)
A. Student Characterisitcs				
Fraction Male	0.553	0.377	0.550	0.370
Fraction White	0.551	0.149	0.671	0.209
Fraction Hispanic	0.348	0.618	0.189	0.524
Fraction Black	0.031	0.222	0.033	0.254
Fraction Other Race/Ethnicity	0.070	0.012	0.106	0.013
Fraction Student-Years FRPL	0.092	0.749	0.071	0.776
Fraction Student-Years Dropout-Risk	0.163	0.717	0.120	0.694
B. Student Outcomes				
Standardized Math Score in 6th Grade	0.695	-0.743	0.773	-0.679
Absence Rate in 6th Grade	0.028	0.048	0.026	0.044
Fraction Ever Receive In-School Suspension in 6th Grade	0.060	0.244	0.046	0.225
Ever Ever Receive Out-of-School Suspension in 6th Grade	0.020	0.132	0.013	0.113
Fraction Ever Graduate from HS	0.863	0.589	0.868	0.607
Fraction Ever Enroll in TX College Ages 18-22	0.721	0.315	0.744	0.321
Fraction Appear in Earnings Data	0.871	0.810	0.873	0.797
Average Earnings at Age 27 ($)	33527	15848	34703	16165
Employment Rate at Age 27	0.643	0.588	0.641	0.576
Average Individual Earnings Rank at Age 27	0.559	0.421	0.561	0.422
Number of Student-Years (1,000s)	3678	6073	20028	17512
Number of Students (1,000s)	318	533	1755	1544
Number of Students in HS Grad Sample (1,000s)	203	325	1091	988
Number of Students in College Sample (1,000s)	164	251	874	772
Number of Students in Earnings Sample (1,000s)	100	140	537	428

Notes: This table reproduces the summary statistics in Table I, separately for students who are high- versus low-risk and who attend CIS versus non-CIS schools. High-risk students are those in the top quartile of the case management prediction model described in Section IV and shown in Figure III Panel C. See the notes to Table I for additional details on the construction of these statistics.

TABLE III: Effect of CIS for Low- vs. High-Risk Students

	First Stage	Std. Math Scores	True HS Grad.	2-Yr Coll.	Pred. HS Grad
	(1)	(2)	(3)	(4)	(5)
$\beta_{post,3-5}$: Low-Risk	1.745	0.011	-1.082	-3.162	0.301
	(0.038)	(0.024)	(0.675)	(1.854)	(0.403)
$\beta_{post,3-5} \times HighRisk$: Difference for High-Risk	-0.046	0.094	4.170	5.815	2.062
	(0.034)	(0.029)	(1.189)	(2.387)	(0.617)
Observations (1,000s)	784	703	745	272	784

Notes: This table presents reduced form DID coefficients from single regressions, estimating the effect of CIS for low- and high-risk students separately on several key academic outcomes. We estimate the regressions using a modified version of equation 2 on the sample of students who are defined either as low- or high-risk. Concretely, we estimate:

$$\begin{aligned} y_{i,t} = {} & \beta_{post,1-2}(\mathbf{1}[1 \leq \tau \leq 2] \cdot CIS_s) + \beta_{post,3-5}(\mathbf{1}[3 \leq \tau \leq 5] \cdot CIS_s) \\ & + \beta_{post,0-2}(\mathbf{1}[1 \leq \tau \leq 2] \cdot CIS_s) \times HighRisk_i + \beta_{post,3-5}(\mathbf{1}[3 \leq \tau \leq 5] \cdot CIS_s) \times HighRisk_i \\ & + \tilde{\lambda}_{s,T(s),r,m,HR} + \tilde{\gamma}_{t,T(s),HR} + \varepsilon_{i,t}, \end{aligned}$$

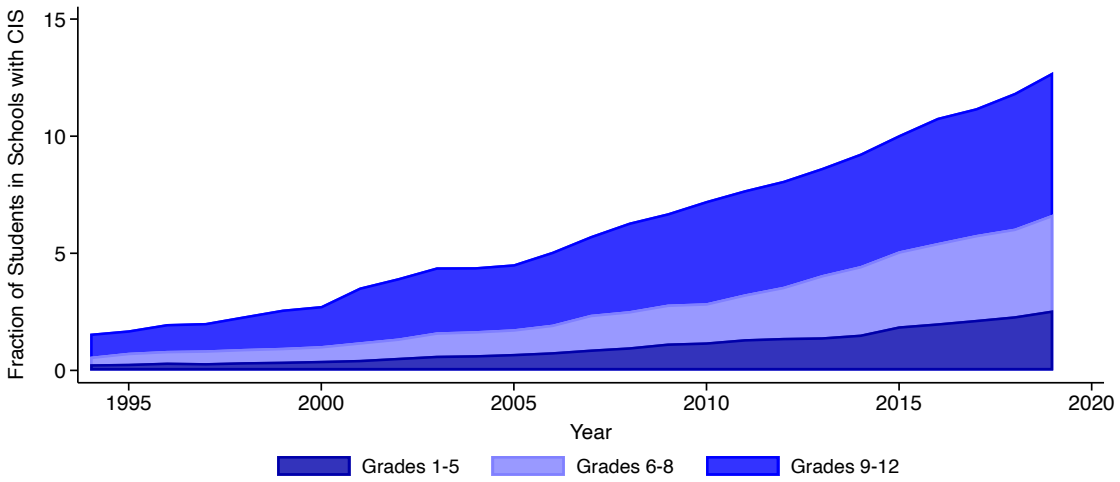
where we interact the DID coefficients and the λ and γ controls with an indicator for being in the high risk group. We also report the mean value of the outcome variable for individuals in the control group in $t = -1$. Standard errors are clustered at the school-by-birth-cohort level. See Appendix Table A.3 for the analogous table with standard errors clustered at the school level.

TABLE IV: Effect of Three Years of CIS on Educational Outcomes for High- and Low-Risk Students

	Std. Math Test Score (1)	HS Graduation (2)	2-Yr Coll. Enrollment (3)
A. High-Risk Students			
3 Years of CIS	0.183	3.445	2.881
	(0.039)	(1.158)	(1.931)
Observations (1,000s)	417	429	161
B. Low-Risk Students			
3 Years of CIS	0.020	-1.103	-2.915
	(0.042)	(0.696)	(1.681)
Observations (1,000s)	286	316	111

Notes: This table presents instrumental variables estimates on the effect of three years of exposure to CIS on educational outcomes, separately for high- and low-risk students. Column (1) presents the effect of three additional years of CIS in middle school on standardized math scores, related to the first stage and reduced form specifications presented in Figures IV and V, respectively. Columns (2) and (3) present the effect of three additional years of CIS exposure over any points, related to the first stage and reduced forms shown in Appendix Figure A.7 Panel D and Figure VIII, respectively. Standard errors are clustered at the school-by-birth-cohort level.

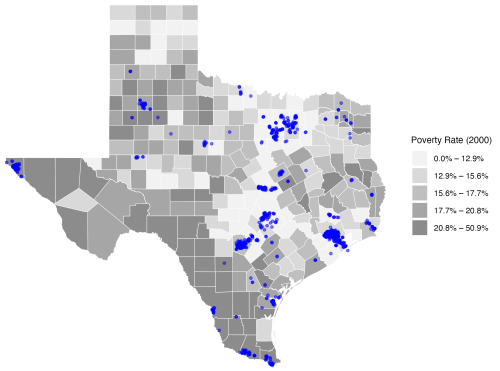
FIGURE I: CIS Program Rollout Across Texas



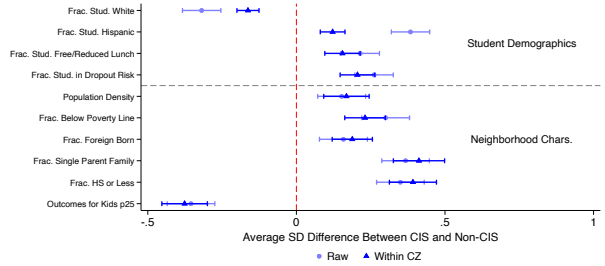
Notes: This figure plots the share of all Texas students enrolled in schools with a CIS program from 1994 to 2019. For each year, we identify the schools with an active CIS program and calculate the number of students enrolled in those schools divided by the total statewide student enrollment. We decompose this total by school level—elementary (grades 1–5), middle (grades 6–8), and high school (grades 9–12)—based on each school's predominant grade span. The shaded regions show the fraction of all Texas students enrolled in CIS elementary, middle, and high schools, respectively, so that the vertical height of each band reflects the share of entire the statewide student population served by CIS at that level and the total height reflects the fraction of all Texas public school students in a school with a CIS program in that year.

FIGURE II: CIS Location Characteristics

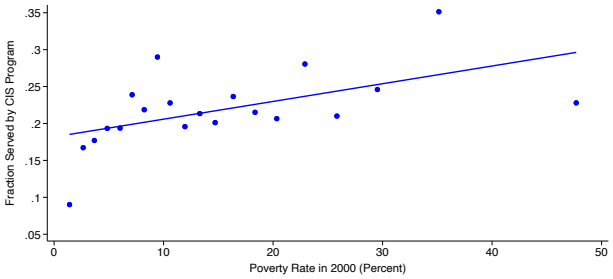
A: Map of CIS Schools in Texas and County-Level Poverty Rates



B: School and Neighborhood Correlates of CIS Programs



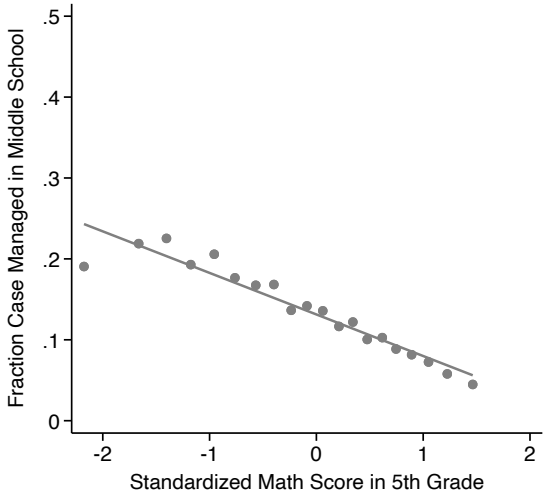
C: Relationship Between CIS Presence and Census Tract-Level Poverty Rates



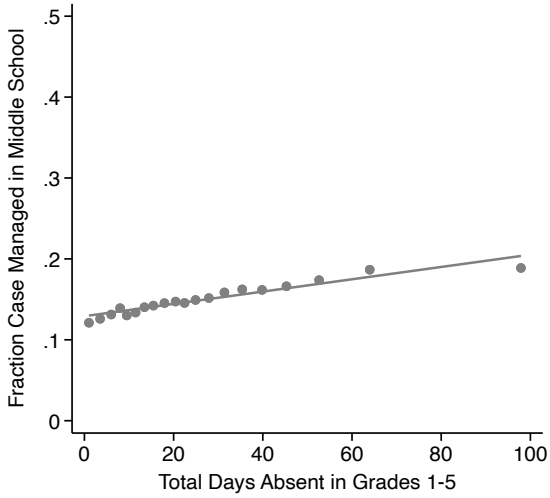
Notes: This figure presents descriptive evidence on the locations and characteristics of CIS programs in Texas. Panel A overlays the locations of CIS programs as of 2019 onto a county-level map of Texas, with counties shaded according to poverty rates from the 2000 Census. Panel B plots differences in school and neighborhood characteristics between schools with and without CIS programs. The light blue series (circles) shows raw differences in standard deviation units, while the dark blue series (triangles) restricts comparisons to schools within the same commuting zone. Student demographic characteristics are measured in the using TEA administrative data described in Section IVA. Neighborhood characteristics are drawn from the 2000 Census and assigned to schools using school catchment zones following Chetty et al. (2025*a*). Panel C presents a binned scatterplot of the relationship between 2000 Census tract poverty rates and CIS program coverage in 2019. To construct the figure, we divide Census tracts into 20 equal population-sized bins by poverty rate and plot the fraction of tracts in each bin that are served by a CIS program against the average poverty rate in the bin, along with a line of best fit.

FIGURE III: Predicting CIS Case Management

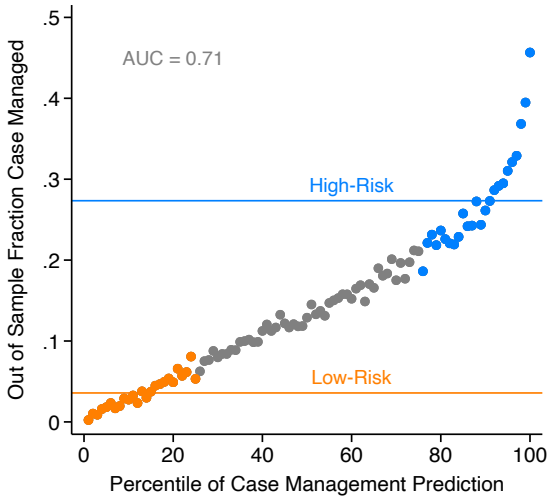
A: Relationship with Test Scores



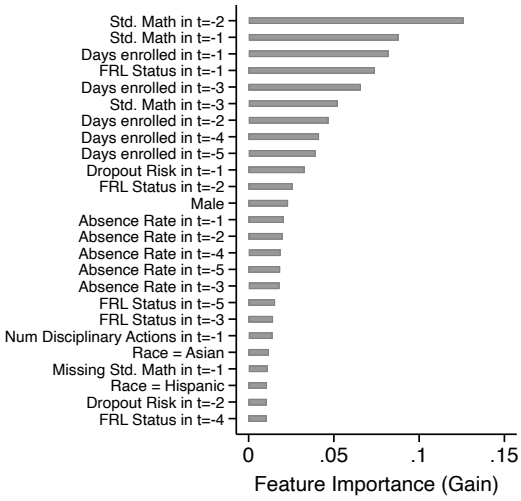
B: Relationship with Attendance



C: Out-of-Sample Accuracy of Case Management Prediction

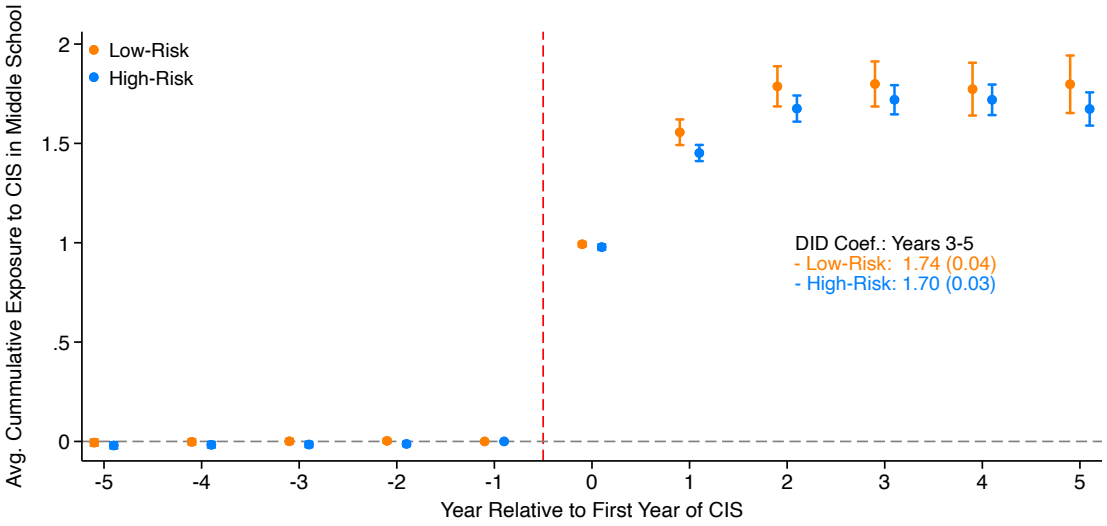


D: Feature Importance – Model Gain



Notes: This figure summarizes the empirical prediction of CIS case management in middle school using pre-treatment student characteristics measured in elementary school. The plots are constructed using a 20% hold-out test set not included in model training. Panel A shows a binned scatterplot of the relationship between standardized 5th grade math scores and the probability of being case managed. Panel B shows the analogous relationship using cumulative absences in grades 1–5. Panel C reports a binned scatterplot of out-of-sample rates of case management by predicted percentile of predicted case management status, based on a gradient boosting model trained on student elementary school outcomes. We classify the top quartile of the predicted distribution as high-risk and the bottom as low-risk and refer to these groups throughout. Panel D presents the top 25 predictors in the GBM ranked by model gain. Predictor variables are measured in years relative to when students first enroll in middle school, ranging from one year prior ($t = -1$) to five years prior ($t = -5$). The model is estimated using data from middle schools with active CIS programs between 2014 and 2019. The dependent variable is an indicator for ever being case managed, as defined in Section III. B, and predictors include student demographics, test scores, attendance, and disciplinary history. See Section IV and Appendix B for additional details on the prediction model.

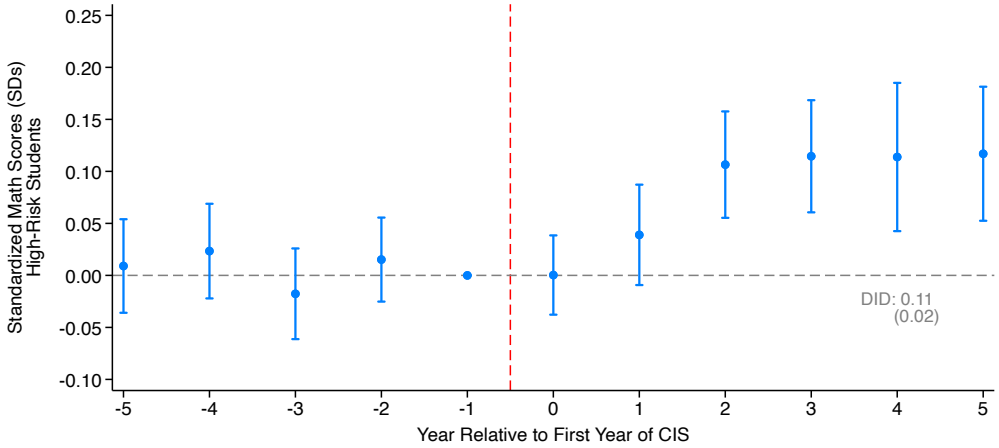
FIGURE IV: First Stage: Exposure to CIS in Middle School



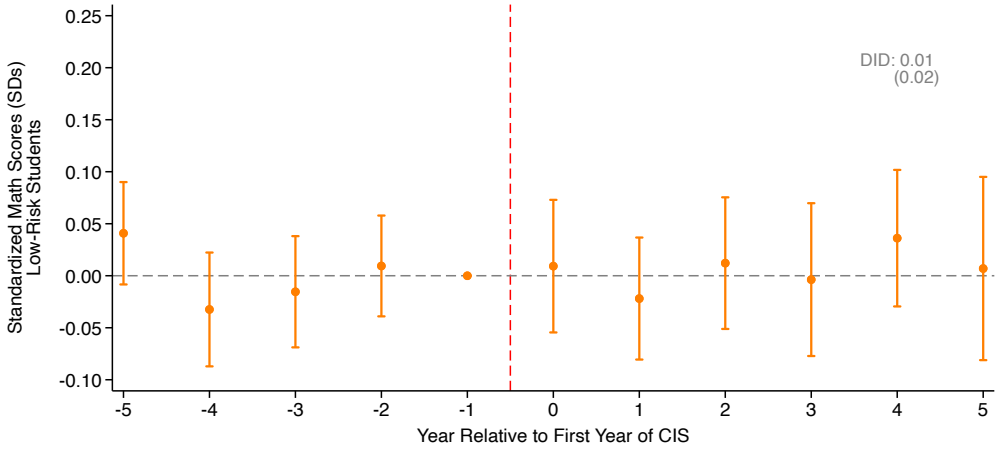
Notes: This figure plots the first stage relationship between time relative to the start of a CIS program and cumulative exposure to CIS in middle school, separately for high-risk (blue) and low-risk (orange) students. See Section IV and the notes to Figure III for details on the assignment to groups. For each group, we estimate Equation 1, comparing treated schools to matched control schools, where the dependent variable is the number of years a student has been exposed to CIS in middle school as of year t. The figure presents event study coefficients estimated using data on students in middle school between 1998 to 2016. In addition to the event study, we estimate a simplified difference-in-differences specification (Equation 2) and report the average difference in CIS exposure between treated and control students over relative years 3 to 5, compared to the pre-treatment period ($\beta_{post,3-5}$). In Table III we report the DID coefficients from a single, fully interacted regression including both high- and low-risk students. Standard errors are clustered at the school-by-birth-cohort level.

FIGURE V: Reduced Form Effects of CIS on Standardized Math Scores

A: High-Risk Students



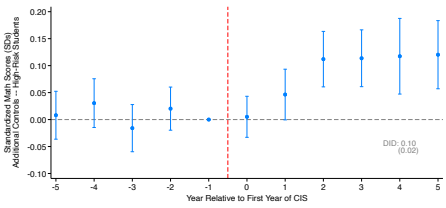
B: Low-Risk Students



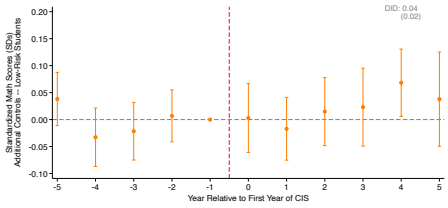
Notes: This figure presents reduced-form estimates of the effect of CIS on standardized math scores in middle school, separately for high-risk (Panel A) and low-risk (Panel B) students. See Section IV and the notes to Figure III for details on the assignment to groups. For each group, we estimate Equation 1, comparing treated schools to matched control schools, where the outcome is the standardized math score of students in year t, standardized within grade and calendar year. The figure shows event study coefficients estimated using data on middle school standardized test taken between 1998 and 2016. In addition to the event study, we estimate a simplified difference-in-differences specification (Equation 2) and report the average effect on test scores over relative years 3 to 5, compared to the pre-treatment period ($\beta_{post,3-5}$). Difference-in-difference coefficients for the other risk quartiles (medium-high and medium-low) as well as students with missing predictions are reported in Appendix Figure A.2. Standard errors are clustered at the school-by-birth-cohort level.

FIGURE VI: Robustness of Baseline Treatment Effects for Std. Math Scores

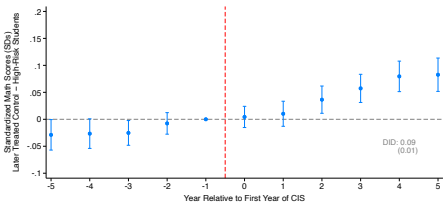
A: Flexible Time Controls by Race and Sex
High-Risk



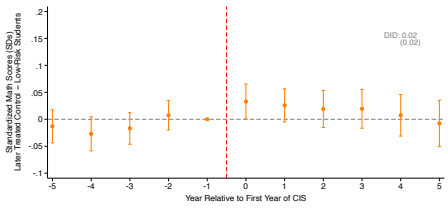
B: Flexible Time Controls by Race and Sex
Low-Risk



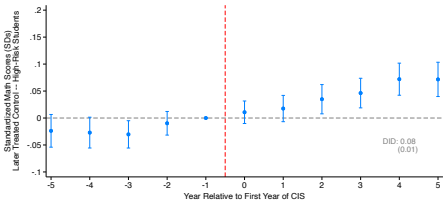
C: Alternate Design: Later Treated Schools
High-Risk



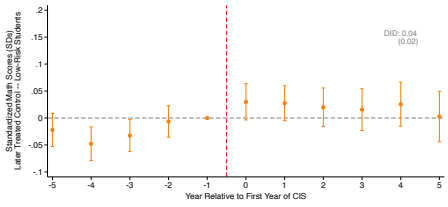
D: Alternate Design: Later Treated Schools
Low-Risk



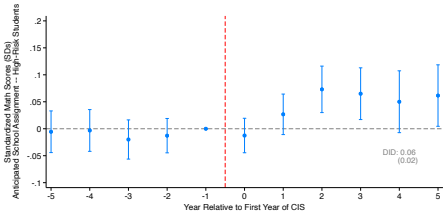
E: Within-CZ Comparison
High-Risk



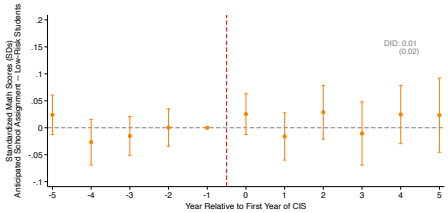
F: Within-CZ Comparison
Low-Risk



G: Assign to First School
High-Risk



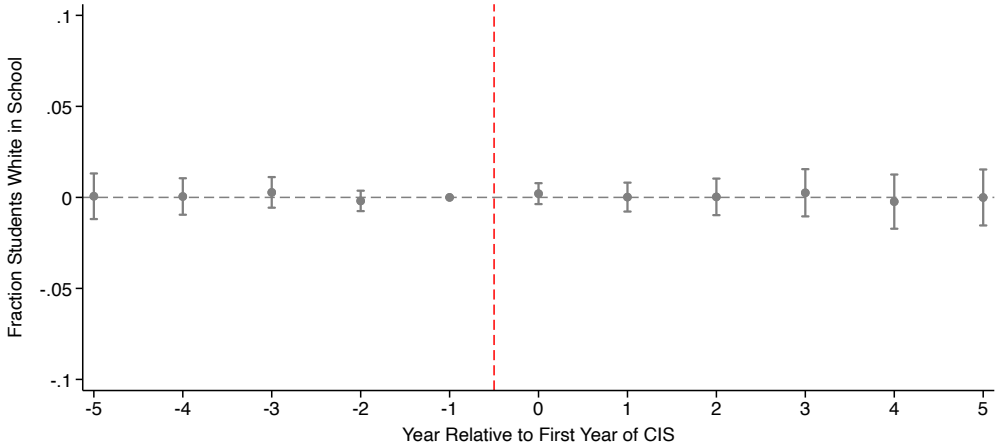
H: Assign to First School
Low-Risk



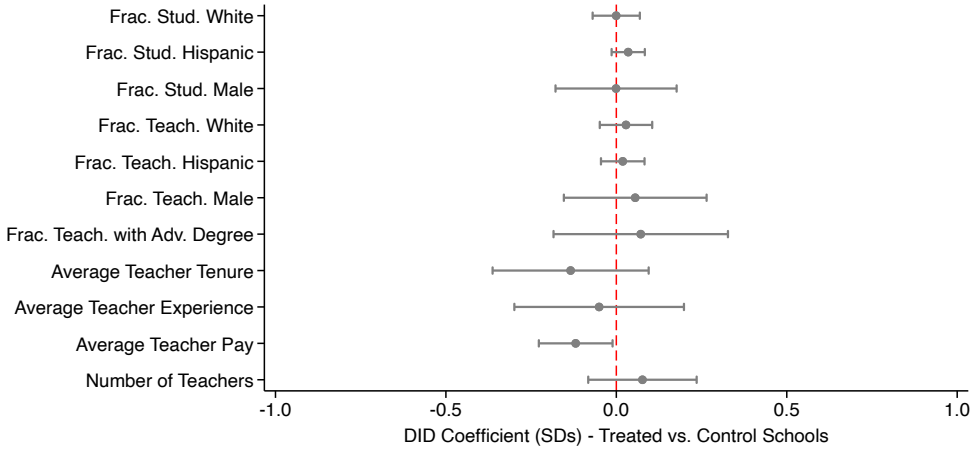
Notes: This figure presents a series of robustness checks for the baseline results shown in Figure V. Panels A and B modify the specification in Equation 1 to interact the γ year fixed effects with race/ethnicity and sex indicators, separately for high- and low-risk students. Panels C (high-risk) and D (low-risk) are estimated using a staggered adoption design in which later-treated CIS schools serve as the control group for earlier-treated schools. For each implementation cohort $T(s)$ and relative time τ, the control group is defined as the set of schools s' such that $T(s') > T(s) + \tau$. See Section V.A for further details. Panels E (high-risk) and F (low-risk) extend this staggered design by interacting the γ year fixed effects with commuting zone fixed effects, ensuring comparisons occur within commuting zones. Panels G (high-risk) and H (low-risk) return to the matched-control design but assign students to a predicted middle school based on their first elementary school enrollment. To do so, we construct a school-to-school transition matrix by grade and birth cohort and assign students to the modal transition path. We then re-estimate Equation 1 among students predicted to appear in the treatment and control schools. See Section IV.A for additional details and Appendix Figure A.4 for the IV estimates accounting for differences in the first stage due to moves during childhood. In all panels, standard errors are clustered at the school-by-birth-cohort level.

FIGURE VII: Change in Covariates After CIS Program Begins

A: Fraction of Students White



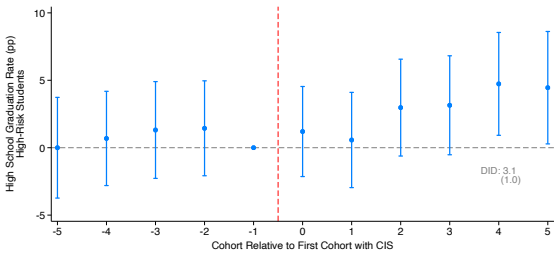
B: DID Coefficients for Student and Teacher Characteristics



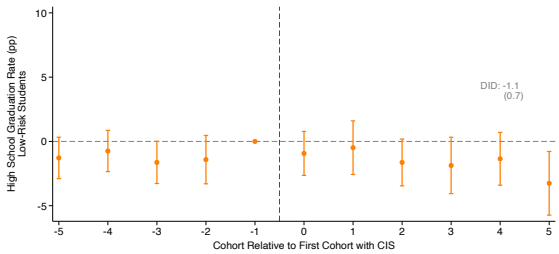
Notes: This figure assesses whether student and teacher characteristics change in treatment schools relative to matched control schools following the start of a CIS program. Panel A estimates an event study version of Equation 1, omitting race/ethnicity and sex interactions since the dependent variable—the fraction of students who are white—is not group-specific. Panel B presents results from a difference-in-differences specification (Equation 2), again excluding demographic interactions, applied to a range of student and teacher covariates. Coefficients are scaled in standard deviation units for comparability, and we report estimates of $\beta_{post,3-5}$, representing average post-treatment differences in years 3 to 5. Standard errors are clustered at the school level.

FIGURE VIII: Effect of CIS on Long-Run Educational Outcomes

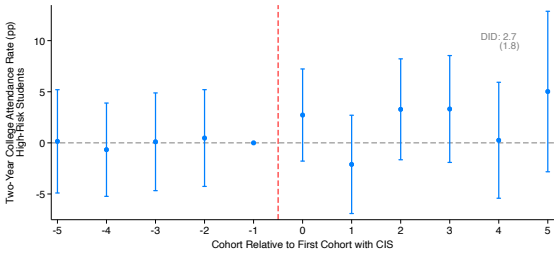
A: High School Graduation
High-Risk



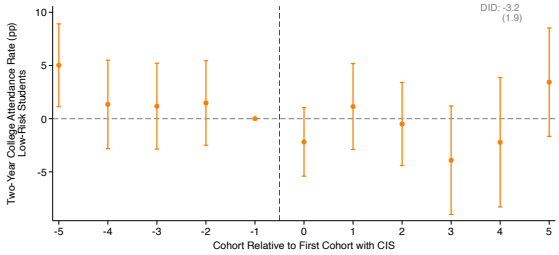
B: High School Graduation
Low-Risk



C: Two-Year College Attendance
High-Risk



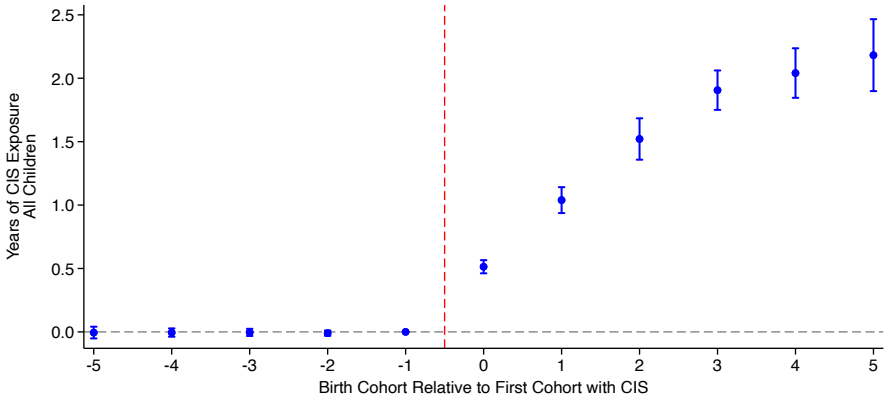
D: Two-Year College Attendance
Low-Risk



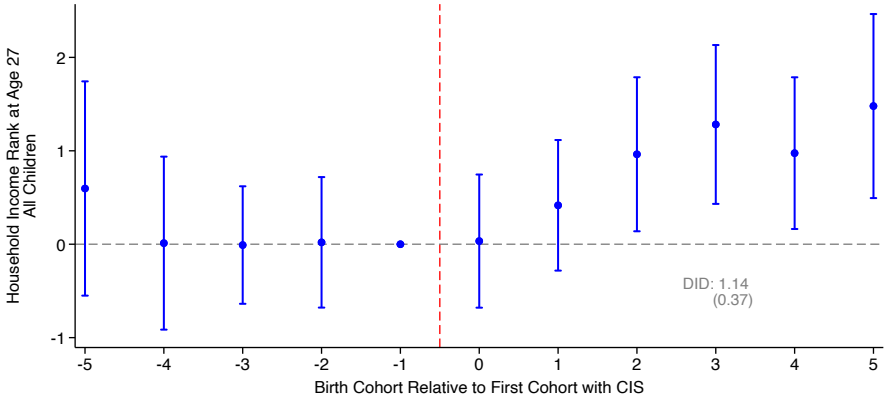
Notes: This figure presents event study coefficients estimating the effect of CIS on long-run student outcomes. Panels A and B show effects on high school graduation rates for high- and low-risk students, respectively, using data on students in middle schools between 1998 to 2016, and estimated using Equation 3. High school graduation is defined as having graduated by 2022. Panels C and D estimate effects on 2-year college attendance, again separately by risk group using data on students in middle school between 1998 and 2009. We define college attendance as an indicator for ever enrolling in a 2-year college (in Texas) between ages 18 and 22 by 2019. Standard errors are clustered at the school-by-birth-cohort level.

FIGURE IX: First Stage and Reduced Form Effect of CIS Household Income Rank at Age 27
Census-IRS Sample: All Children

A: First Stage

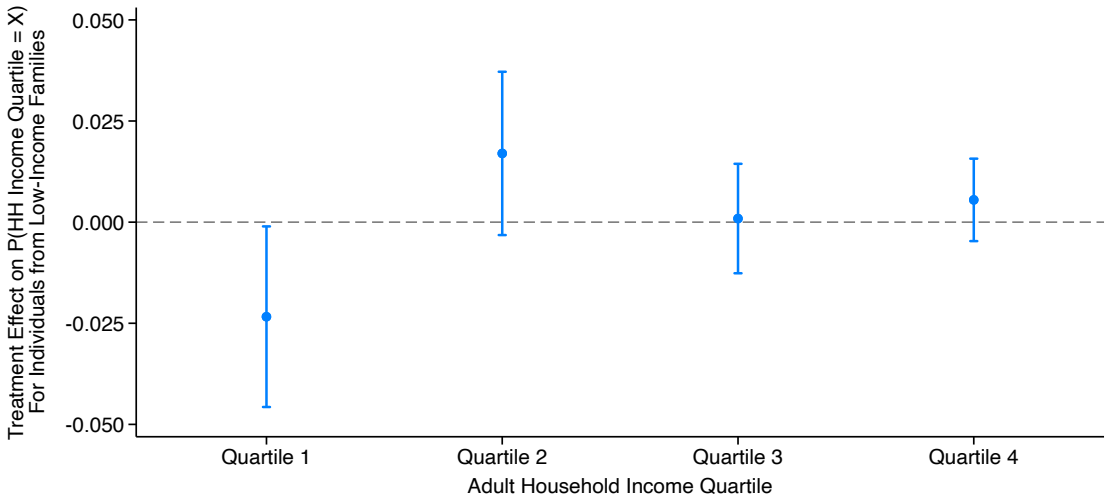


B: Reduced Form



Notes: This Figure presents the first stage and reduced form effects of CIS using the Census-IRS data described in Section III C. In both panels, we plot event study coefficients estimated from Equation 4. In Panel A, the dependent variable is the average number of years of exposure to a CIS program. In Panel B, the dependent variable is adult household income rank measured at age 27. Standard errors are clustered at the high school catchment zone boundary level. Census Bureau Disclosure Review Board no. CBDRB-FY23-CES014-028 (Goldman, Gracie and Porter, 2023).

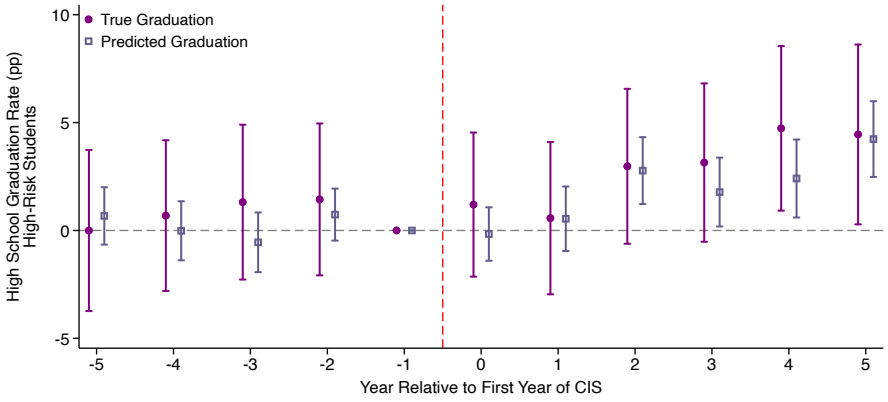
FIGURE X: Effect of CIS Exposure on Adult Income Quartiles
Census-IRS Sample: Low-Income Children



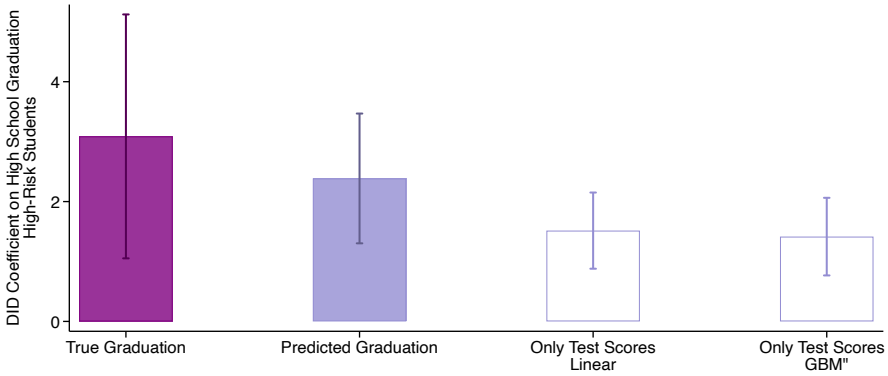
Notes: This figure shows the effect of CIS on the probability that low-income students fall into different quartiles of the income distribution at age 27. It is estimated using Census-IRS data, described in Section III.C. We show difference in difference coefficients from four separate regressions, where in each the dependent variable is an indicator equal to 1 if a student is in quartile X at age 27. Coefficients capture average effects for students born 3–5 years after the implementation cohort, estimated using the difference-in-differences analog of Equation 4. Standard errors are clustered at the high school catchment zone level. Census Bureau Disclosure Review Board no. CBDRB-FY23-CES014-028 (Goldman, Gracie and Porter, 2023).

FIGURE XI: Comparing True vs. Predicted Long-Run Outcomes

A: High School Graduation
High-Risk Students



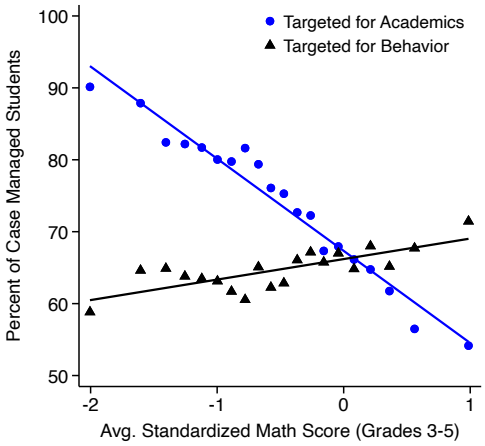
B: DID High School Graduation
High-Risk Students



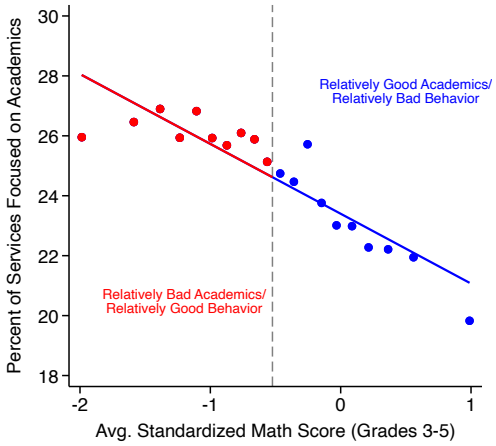
Notes: This figure compares the true effect of CIS on high school graduation for high-risk students to predicted effects based on contemporaneous outcomes. In Panel A, we present event study coefficients estimated using the matched control design. The dark purple series (circles) replicates the result shown in Figure VIII Panel *A*. The lighter series (squares) shows coefficients from Equation 1, where the dependent variable is predicted high school graduation based on a gradient boosting model trained on contemporaneous academic and behavioral outcomes. See Appendix Figure A.12 for details on the prediction model. Panel B reports difference-in-differences estimates from each specification. Hollow bars show two alternative predictions based only on test scores. In the first, we use a linear regression of graduation on middle school test scores, estimated using statewide data from 1998–2004. In the second, we fit a flexible gradient boosting model on the same data to capture possible nonlinear relationships. Standard errors are clustered at the school-by-birth-cohort level.

FIGURE XII: The Role of Personalization

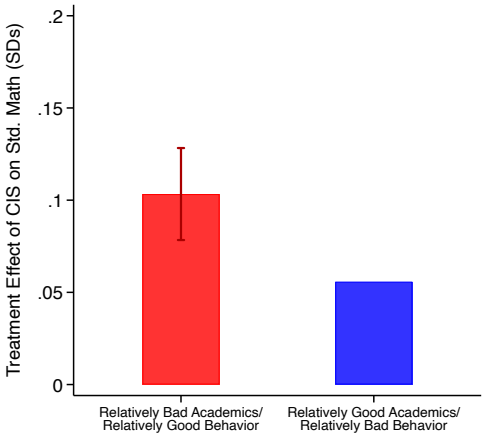
A: Targeted Areas of Support Among Case-Managed Students



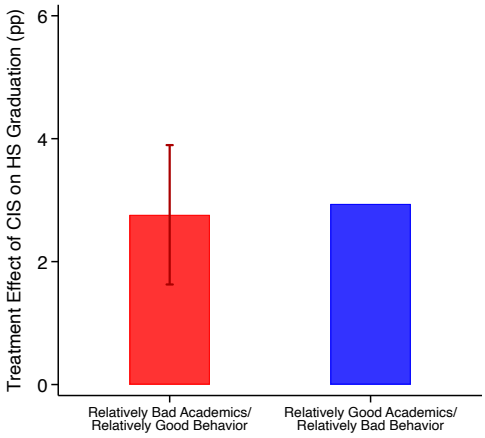
B: Percent of Services Academic Among Case-Managed Students



C: Effect of CIS on Std. Math Scores By ES Math Performance



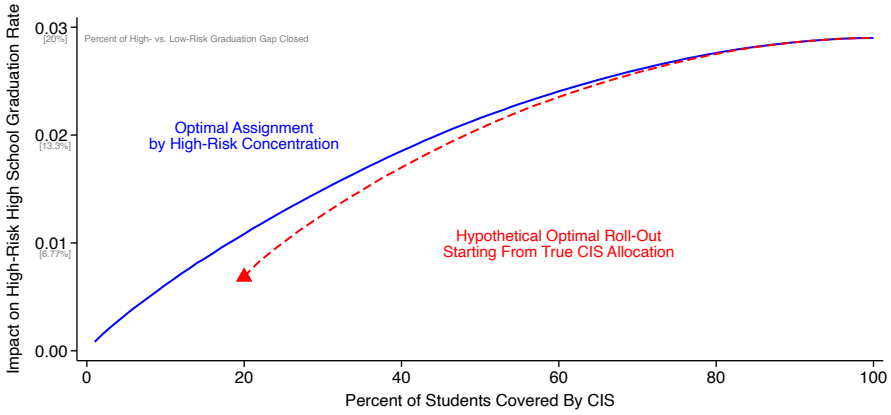
D: Effect of CIS on HS Graduation By ES Math Performance



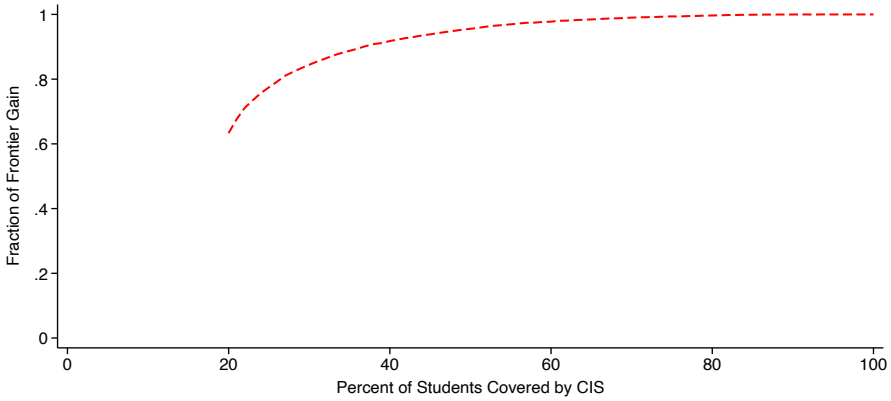
Notes: This figure presents results on the personalization of CIS services. Panel A shows the relationship between the fraction of students in case-management flagged for academics and behavior and their average standardized math scores in grades 3-5. Panel B plots the relationship between the fraction of academic services a student receives and their average standardized math scores in grades 3–5. In both panels, the sample includes case-managed sixth graders between 2016 and 2019 from the CIS microdata, excluding the 20 percent of students with the fewest navigator interactions. In Panel A we remove school-year observations where 100% of students are flagged for academics. The figures present a binned scatterplot controlling for students' predicted probability of case management and year fixed effects. Students can be flagged for any combination Academics, Behavior, Attendance, and/or Social Services (see Section III.A for details and Appendix Table A.1 for the full breakdown). We define a student as flagged for academics (behavior) if academics (behavior) is included in their list of targeted areas of support. Academic services are defined in Section VI. Panels C (test scores) and D (high school graduation) show difference-in-differences (DID) estimates from a modified version of Equation 2, corresponding to the specification in Figure V Panel *A*. The specification is estimated at the individual level and includes controls for predicted case-management risk, an indicator for below-median elementary-school test scores, and interactions between this indicator and the two post-period × treatment indicators. The blue bars (right) report the estimated treatment effects for students with above-median prior test scores, while the red bars (left) report the corresponding effects for students with below-median scores. Vertical lines show 95 percent confidence intervals for the difference between the two group-specific treatment effects. Standard errors are clustered at the school-by-birth-cohort level.

FIGURE XIII: CIS Expansion Counterfactual

A: Frontier Allocation vs. Current CIS Allocation



B: Hypothetical CIS Roll-Out: Fraction of Frontier Gain



Notes: This figure presents the results of a simulation estimating the effects of a hypothetical expansion of CIS programs. Motivated by the results in Figure VIII, where gains are concentrated among high-risk students, the blue series in Panel A simulates a counterfactual rollout in which CIS is expanded to schools in descending order of high-risk student concentration. The horizontal axis plots the share of all middle school students covered by CIS; the vertical axis shows the corresponding increase in the statewide high-risk high school graduation rate (equal to the treatment effect of CIS times the fraction of high-risk students served by the program. The triangle marks the actual 2019 CIS allocation, which covers approximately 20% of middle school students. The vertical gap between the actual allocation and the blue frontier curve reflects that CIS programs are not strictly allocated to the highest-need schools (as measured by the amount of high-risk students in the school). The red dashed series shows a hypothetical expansion beginning from the existing CIS allocation and continuing to schools with the next-highest high-risk shares. Panel B shows the fraction of the frontier gain (from Panel A) achieved under this sequential rollout. The high versus low-risk gap in Panel A is based on a 15 percentage point gap in graduation rates today among schools that have never had a CIS program.

Supplementary Appendix

A Additional Details on Data Construction

Our analysis primarily uses administrative data from the Texas Education Agency (TEA), supplemented with postsecondary enrollment data from the Texas Higher Education Coordinating Board (THECB) and quarterly earnings records from the Texas Workforce Commission (TWC).

We begin by assembling a student-level enrollment panel starting in 1994. Each year's enrollment file reports the student's campus of attendance; basic demographics (sex, race/ethnicity, and age as of September 1); and indicators for free- or reduced-price lunch eligibility and "at risk of dropping out" status. After constructing the panel, we clean invariant demographics (sex, birth cohort, and race/ethnicity) to enforce within-student consistency. When a student's reported value varies across years—which is rare, with fewer than 1% of students having inconsistent sex or cohort information and about 4% having inconsistent race/ethnicity—we assign the modal value.[1] Because enrollment data are collected at a single point in the fall semester, it is also uncommon for a student to appear in more than one campus within a year: fewer than 1% of students are listed at multiple schools in the same academic year. In such cases, we randomly assign the student to a single campus for that year. The resulting enrollment panel serves as the spine for all data linkages used in the analysis.

Our primary contemporaneous outcome is students' performance on standardized tests in grades 3–8. Over our study period, Texas administered three different assessments: the Texas Assessment of Academic Skills (TAAS, 1990–2002), the Texas Assessment of Knowledge and Skills (TAKS, 2003–2012), and the State of Texas Assessments of Academic Readiness (STAAR, 2012–present). To ensure comparability across regimes, we standardize all scores within exam–grade–year cells and include year fixed effects in all specifications.

The administrative testing files sometimes contain multiple records for the same student in the same exam–grade–year because Texas reports scores from the initial administration as well as make-up tests and retakes. When multiple scores are present, we assign the student the maximum score. This choice reflects the institutional purpose of retakes—students may demonstrate mastery following an absence or a low initial score—and reduces noise arising from within-student variation. For instance, in the 2009 8th-grade TAKS math exam, 20% of students have multiple records, some of which contain scores of 0 indicating a retake. When students have multiple non-zero or non-missing scores, they tend to be close in value. The within-student variance among these duplicates (4.1 points) is less than half the overall variance in that exam–grade–year (9.1

[1] Under the Education Research Center's FERPA review process, we are prohibited from reporting percentages below 1%, regardless of sample size.

points). Using the maximum score therefore provides a consistent measure of each student's demonstrated achievement without materially altering the score distribution.

In addition to test scores, we use detailed attendance records reported every six weeks. For each student–campus–grade–period cell, the data include the number of instructional days the student was enrolled and the number of days absent. It is extremely uncommon for a student to have multiple records within the same student–campus–grade–period; when this occurs, we retain the record with the highest number of instructional days and break any remaining ties randomly.

A small share of cases, however, reflect students appearing at multiple campuses within the same six-week period. In 2015, 1.3% of student–period cells contain more than one observation, and 5.9% of students experience at least one such period. In these cases, we assign the student the average number of days enrolled and absent across the duplicates. These multi-campus records often arise when students change schools mid-period, and it is not possible to determine which campus's record most accurately reflects the true number of days enrolled or absent. Averaging across duplicates prevents movers from being assigned spuriously extreme values—such as a full set of absences at one campus when the student had already transferred—which could otherwise mechanically inflate absence measures.

Our final set of contemporaneous outcomes comes from student-level disciplinary records. These data are reported at the student-by-incident level and include the type of action (in-school suspension, out-of-school suspension, or expulsion) and, for suspensions, the number of days assigned. We aggregate these records to the student-by-year level, constructing measures of the total number of disciplinary actions (overall and by type), the total days assigned, and indicators for ever receiving each type of action in a given year.

Finally, we use TEA administrative data to measure high school graduation as a long-run outcome. The graduation files list all students who graduate in a given year and the campus from which they graduate. Our primary measure is an indicator equal to 1 if a student ever appears in the graduation records. We observe graduation outcomes through 2022 and code students who never appear in these files as non-graduates. Our findings are robust to alternative definitions, including graduating by age 18, by age 20, or at any point between ages 16 and 20. It is extremely rare for a student to have multiple graduation records (<1%), but when this occurs, we assign the first observed graduation year when constructing age-based measures.

We link TEA records to college enrollment data from THECB using ERC-provided student identifiers. The THECB data cover enrollment at Texas higher education institutions from 1992 to 2019. Enrollment is observed separately in the fall, spring, and summer terms, and institutions are categorized into four groups: 2-year colleges, 4-year colleges, health institutions, and independent institutions. Our analysis focuses on 2-year and 4-year colleges, which represent the public higher education sector in Texas. We define college attendance as ever enrolling in a given sector (2-year

or 4-year) between ages 18 and 22. Out-of-state enrollment is not observed for most of our sample; National Student Clearinghouse data are only available in the most recent years. As a result, students who attend college exclusively out of state cannot be distinguished from non-attenders and are coded as not attending. However, Texas has an unusually high rate of in-state enrollment: in 2020, 84% of first-time college enrollees from Texas remained in-state, roughly 10 percentage points above the national average (U.S. Department of Education, 2022).

Finally, we use earnings data from the Texas Workforce Commission (TWC) to validate the earnings results from the national sample shown in Section V.C. The TWC data consist of quarterly unemployment-insurance (UI) wage records for all workers in Texas covered by UI. These records include worker–employer–quarter observations of earnings; coverage excludes the self-employed, independent contractors, military personnel, and informal-sector workers (Stevens, 2007). We aggregate earnings to the worker-year level by summing across all employers and inflate all dollar values to 2015 dollars using the CPI-U. Because the UI data cannot distinguish between unemployment, employment in non-covered jobs, and out-migration from Texas, we follow standard practice and classify individuals as "active workers" if they ever have positive UI-covered earnings between ages 18 and 25. We impute zero earnings only in years with missing earnings data for this active-worker sample. Using these annual earnings, we construct within-cohort earnings ranks at each age following Chetty et al. (2025*b*).

After merging all data sources, our final dataset contains a set of invariant student characteristics (sex, birth cohort, race/ethnicity); year-varying characteristics (campus attended, free- or reduced-price lunch status, and dropout-risk indicators); year-specific contemporaneous outcomes (test scores, attendance, and disciplinary actions); and student-invariant long-run outcomes (high school graduation, college attendance, and earnings).

We use data provided directly by CIS to identify which schools operated CIS programs and the year each program began. The original file lists 1,216 CIS schools as of 2019–20. We first exclude 38 records with unusable NCES school IDs (missing, duplicated, or not corresponding to Texas schools). We then drop 38 schools classified as "alternative schools" (e.g., alternative education centers or career centers). An additional 101 schools lack information on the program start date, and 27 schools have NCES IDs that do not match to a TEA campus code, preventing linkage to the microdata. After applying these restrictions, we are left with 1,012 usable CIS program records.

To construct the analysis sample, as described in Section IV, we focus on middle schools and match CIS-treated schools to never-treated controls. For each CIS implementation cohort, we take the set of middle schools receiving a CIS program in a given year (e.g., schools first treated in 2004) and compile school- and student-level data from six years prior to program adoption (1998 in this example). We then append data from all never-treated middle schools in that same baseline year. This forms the matching dataset for the 2004 implementation cohort. We repeat this procedure for

all implementation cohorts from 2001 through 2019.

Using the stacked data, we estimate a regression predicting whether a school adopts a CIS program in year t based on its characteristics in year t-6. The predictors include the fraction of students classified as "at risk of dropping out," the fraction white, the schoolwide absence rate, the fraction of high-risk and low-risk students (see Appendix B for details), and the predicted high school graduation rate based on contemporaneous characteristics (Section VI). We also include implementation-cohort fixed effects. For each school, this model yields a predicted probability of CIS adoption.

For each treated school, we compare its predicted probability to those of all potential control schools in the same implementation cohort, excluding never-treated schools in the same district. We then select the control school with the closest predicted probability of CIS.

B Additional Details on Case Management Prediction

As discussed in Section IV, our analysis focuses on students who would be most and least likely to interact with a CIS site coordinator. This requires constructing predicted probabilities of case management for students in both treated and control schools.

We estimate these predictions using data on actual case-management status between 2014 and 2019. We assemble a dataset containing 103,000 students who were ever case managed in grades 6–8 during this period and 612,000 other students in the same schools and years who were never case managed in middle school. For each student, we construct a set of predictors based on their elementary school characteristics in the five years prior to their middle-school observation.[2]

We construct a series of in-school outcomes from test score, attendance, and disciplinary records. From the testing data, we include standardized math scores and an indicator for missing test scores in a given year.[3] From the attendance records, we include days enrolled and days absent in each six-week period, the annual absence rate, and an indicator for being chronically absent (absent more than 10% of enrolled days). From the disciplinary data, we include the number of in-school and out-of-school suspensions, an indicator for ever receiving either type of suspension, and the total number of days assigned.

Finally, in addition to school-based outcomes, we include a set of demographic predictors. For each student, we use invariant race/ethnicity and sex, as well as year-specific indicators for free- or reduced-price lunch eligibility and TEA "at-risk" status. We do not include birth cohort or school identifiers, as our goal is to generate predictions that can be applied out of sample to

[2] This avoids issues arising from possible grade retention. For example, if a student repeated third grade, we include both years of third-grade data and do not include the year the student was in first grade.

[3] Across these years, not all grades administered standardized tests in every year. For example, in 2015 no STAAR exam was given in first grade.

different schools and years.

Using these data, we estimate the probability of case management with a gradient boosting decision tree algorithm (Ke et al., 2017). We reserve 30% of the sample as a validation set, which is not used in training. The validation data are used solely to assess model performance (e.g., Figure III*C*).

Using the remaining 70% of the data, we train the gradient boosting model. Because CIS case management is a relatively rare outcome—14% of students are case managed in the prediction sample—we combine a standard GBM with a second model that up-weights case-managed observations.[4] We implement this using five-fold cross-validation. For each fold, we proceed as follows:

1. Hold out the fold (20% of the training data);
2. Estimate the basic GBM on the remaining data;
3. Generate predictions for the held-out fold;
4. Estimate the up-weighted GBM on the remaining data;
5. Generate predictions for the held-out fold.

This procedure yields out-of-sample predictions from both models for every observation in the training data. We then regress case-management status on these two sets of predictions using a logit model and save the resulting coefficients.

Finally, we re-estimate each GBM model on the full training sample. This leaves us with two fitted GBM models—one standard and one up-weighted—and the coefficients from the logit model used to blend their predictions. Importantly, because the logit was trained only on out-of-sample predictions generated during cross-validation, its coefficients are not fitted on the same predictions produced by the final GBM models. This prevents overfitting and ensures proper separation between training and blending.

To apply these predictions to the full sample (beyond the training and validation data), we construct an analogous version of the prediction microdata for all years in our dataset. We then apply both GBM models to these back-filled records to generate predicted probabilities from each model. Using the previously estimated logit coefficients, we blend the two sets of predictions to obtain the final predicted likelihood of case management for every student. We classify students in the top quartile of this predicted distribution as high-risk and those in the bottom quartile as low-risk (see Figure III*C*).

[4]This approach follows standard methods for learning from imbalanced data (e.g., He and Garcia, 2009). We up-weight positive cases by a factor of 10.

TABLE A.1: Characteristics of Case-Managed Students

	Mean (1)	p25 (2)	p50 (3)	p75 (4)
A. Student Characterisitcs				
Fraction Male	0.466			
Fraction White	0.128			
Fraction Hispanic	0.667			
Fraction Black	0.179			
Fraction Other Race/Ethnicity	0.026			
Fraction FRPL	0.710			
Fraction Dropout-Risk	0.762			
B. Student Outcomes				
Standardized Math Score in 6th Grade	-0.541	-1.186	-0.708	-0.022
Absence Rate in 6th Grade	0.046	0.012	0.034	0.062
Fraction Ever Receive In-School Suspension in 6th Grade	0.265			
Ever Ever Receive Out-of-School Suspension in 6th Grade	0.146			
C. Reason For Referral				
Referral includes				
Academics	0.834			
Attendance	0.223			
Behavior	0.655			
Social Services	0.347			
D. Information on Interactions				
Days After August 1st Parent Consent Received	77	35	64	126
Number of Interactions in Year	38	18	29	48
Number of Unique Services Received	14	9	13	18
Fraction of Services Individual	0.319	0.019	0.091	0.640
Fraction of Services from External Organization	0.148	0.000	0.000	0.000
Number of Student-Years (1,000s)	170			
Number of Students (1,000s)	132			

Notes: This table presents summary statistics for students who were case managed by CIS in a middle school between 2014 and 2019, using the CIS microdata described in Section III B. Panel A reports student demographic characteristics, and Panel B summarizes student outcomes. Panel C shows the fraction of students referred for each broad category of need (not mutually exclusive). Panel D provides statistics on students' interactions with CIS navigators. A unique service refers to a distinct service code in the CIS microdata. Services are categorized as individual or group-based and are classified by whether they were delivered directly by TEA/CIS staff or by an external provider.

TABLE A.2: Needs Addressed by CIS

	Number of Student-Years (1,000s) (1)	Fraction of Case-Managed Students-Years (2)
Targeted Need Includes		
Test Readiness	70.4	0.410
Grades	70.3	0.409
Academic Readiness	69.6	0.405
Social Skills	61.9	0.360
Self Esteem	55.5	0.323
Basic Needs	53.1	0.309
College Readiness	36.9	0.215
Homework Completion	35.4	0.206
Absences	30.5	0.177
Classroom Conduct	26.3	0.153
Language Development	24.5	0.142
Tardies	18.7	0.109
Classroom Participation	17.9	0.104
Family Conflict	13.2	0.077
Mental Health and Wellness	12.3	0.072
Life Skills	10.4	0.061
Career/Employment	9.2	0.054
Delinquent Conduct	5.8	0.034
Health	3.3	0.019
Grief/Loss	1.9	0.011
Housing	1.4	<0.01
Suspected Substance Use	1.0	<0.01
Violence	0.9	<0.01
Gang Involvement	0.6	<0.01

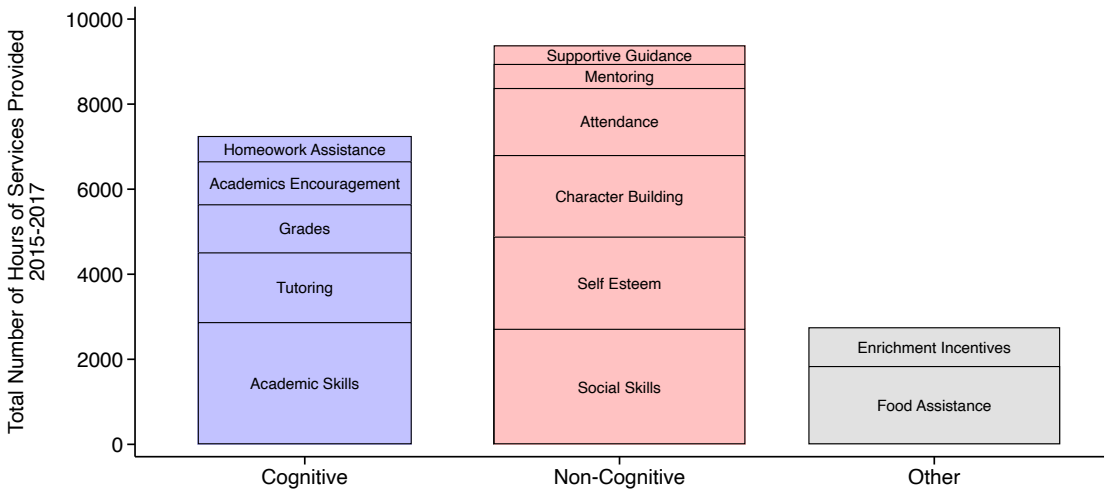
Notes: This table provides information on the detailed needs of students who are case managed by CIS in middle schools from 2014-2019, using the CIS microdata described in Section III B. For each need category, we provide the fraction of student-year observations who are flagged as having this need (not mutually exclusive).

TABLE A.3: Effect of CIS for Low- vs. High-Risk Students

	First Stage	Std. Math Scores	True HS Grad.	2-Yr Coll.	Pred. HS Grad
	(1)	(2)	(3)	(4)	(5)
$\beta_{post,3-5}$: Low-Risk	1.745	0.011	-1.082	-3.162	0.301
	(0.046)	(0.043)	(0.829)	(2.391)	(0.675)
$\beta_{post,3-5} \times HighRisk$: Difference for High-Risk	-0.046	0.094	4.170	5.815	2.062
	(0.040)	(0.049)	(1.331)	(2.810)	(0.845)
Observations (1,000s)	784	703	745	272	784

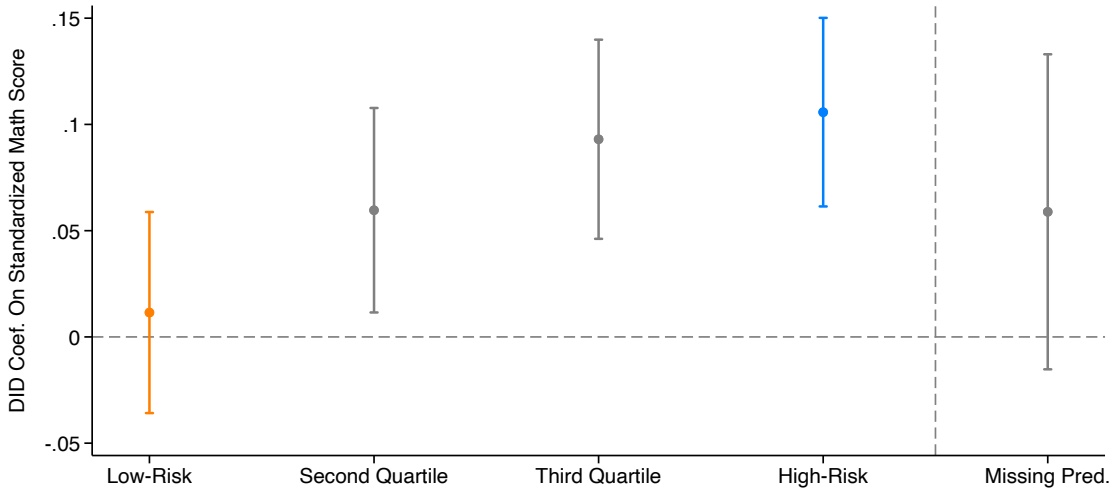
Notes: This table presents results from the same regression as those in Table III, but with standard errors clustered at the school level. See table notes for Table III for additional details.

FIGURE A.1: Top 10 Services Provided by CIS 2015-2017



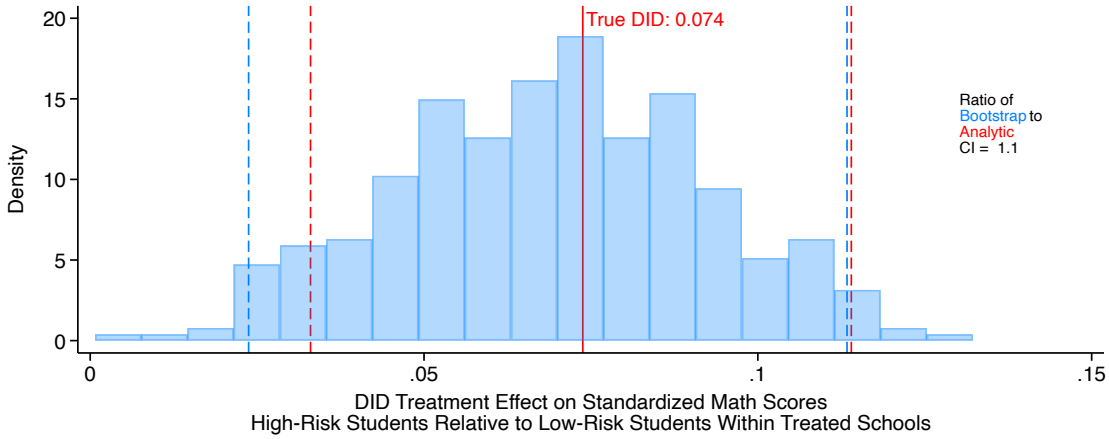
Notes: This figure summarizes the most common services provided by CIS navigators, measured in total time. Using the CIS microdata described in Section III/B, we identify the top 10 most frequently provided services in each year from 2015 to 2017. We classify these services into three broad categories based on our own determination: cognitive, non-cognitive, and other. The height of each bar reflects the total time spent delivering that service across all three years.

FIGURE A.2: Heterogeneity: Effect of CIS on Test Scores By Risk Group



Notes: This figure reports difference-in-differences coefficients estimating the effect of CIS on standardized test scores, separately by student risk status. For each quartile of the risk distribution—as well as for students without a predicted risk score—we report the coefficient $\beta_{post,3-5}$ from Equation 2, comparing treated schools to matched control schools. The estimates for high- and low-risk students correspond to those shown in the event studies in Figure V. Students are missing a risk score if they lack elementary school data. Standard errors are clustered at the school-by-birth-cohort level.

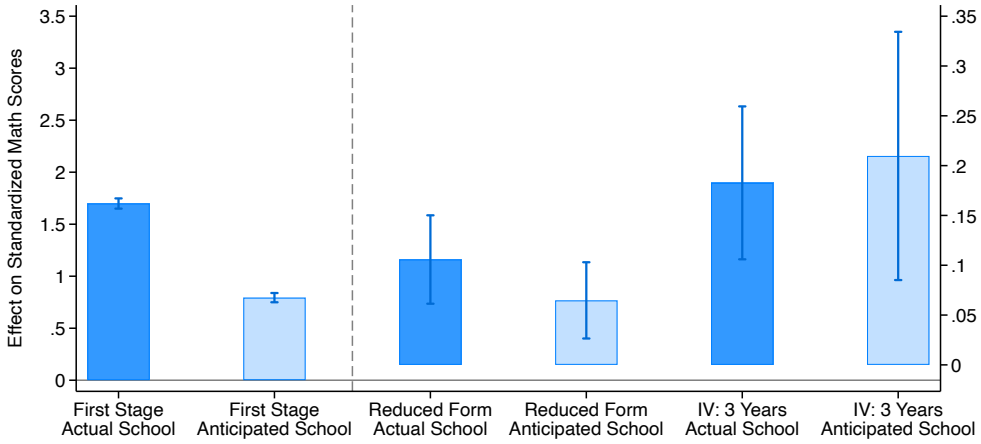
FIGURE A.3: Clustered Bootstrap: Within-School Effects on Test Scores High-Risk Students



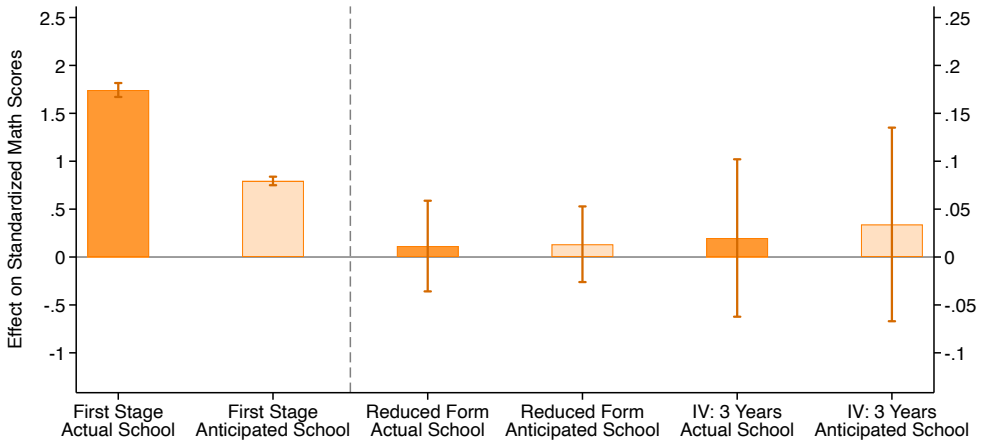
Notes: This figure shows the distribution of treatment effects from a clustered bootstrap procedure. For computational reasons, we do not re-estimate the full prediction model from Figure III or the corresponding treatment effects from Figure V. Instead, we take a simplified approach based on the within-school design shown in Figure A.5, Panel A. Rather than using the original gradient boosting algorithm to assign risk scores, we estimate a logistic regression on the same set of predictors (see Section IV) to classify students into high- and low-risk groups.The bootstrap procedure proceeds as follows: for each iteration, we first resample schools with replacement from the training sample and re-estimate the logistic regression model. We then resample schools with replacement from the analysis sample, use the estimated logistic model to assign risk scores to students in the resampled data, and estimate the within-school treatment effect (shown and explained in Appendix Figure A.5). We repeat this process across bootstrap iterations.The figure shows the resulting distribution of treatment effect estimates. For reference, we also report the treatment effect from the actual data using this simplified design, as well as the ratio of the conventional 95% confidence interval (based on school-clustered standard errors) to the 95% confidence interval implied by the bootstrap distribution.

FIGURE A.4: Effects of CIS on Standardized Test Scores: Actual School Assignment vs. First School Assignment

A: High-Risk Students



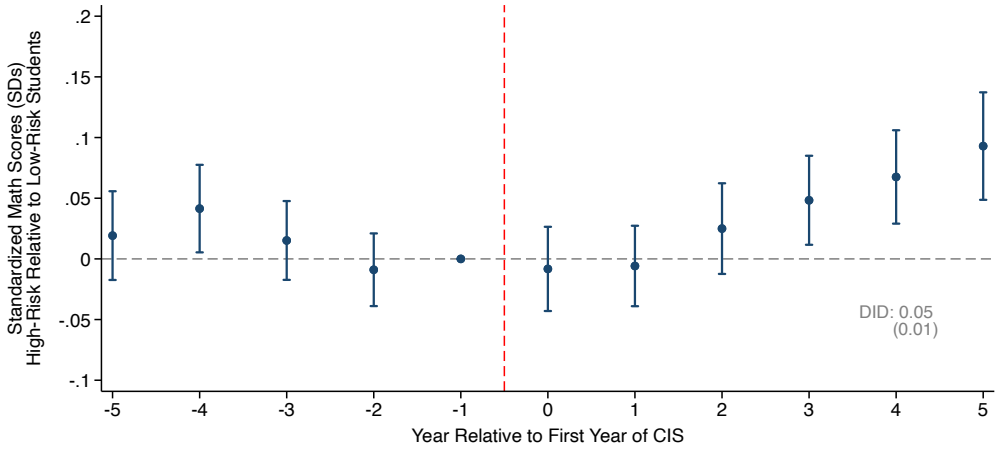
B: Low-Risk Students



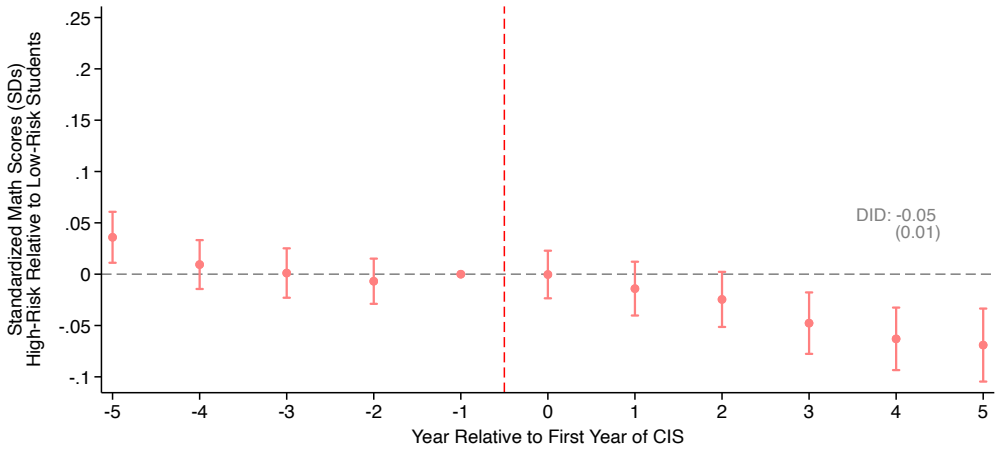
Notes: This figure compares the first stage, reduced form, and IV effects of CIS on standardized test scores using students' actual versus anticipated middle school assignments. In Panel A, we restrict to high-risk students. The first stage and reduced form estimates using actual school assignment replicate those reported in Figures IV and V, respectively. Anticipated school assignment follows the approach used in Figure VI: rather than using students' actual enrolled middle school, we assign them to an anticipated school based on the first elementary school they attend (see notes to Figure VI for details). The reduced form estimate under anticipated assignment is the same as that shown in Figure VI Panel G. The IV estimates rescale the reduced form by the corresponding first stage to yield an effect per three years of CIS exposure. Panel B presents the analogous comparisons for low-risk students. Standard errors are clustered at the school-by-birth-cohort level.

FIGURE A.5: Within-School Effects of CIS on Standardized Test Scores: High-Risk Relative to Low-Risk Students

A: Treated Schools



B: Control Schools



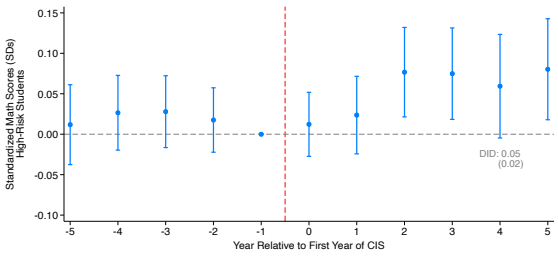
Notes: This figure presents event study coefficients estimating the relative effect of CIS on standardized test scores for high-risk students compared to their low-risk peers within the same school. To do this, we estimate a specification similar to Equation 1, but we modify it to compare outcomes for high-risk versus low-risk students within treated (control) schools. Concretely, we estimate:

$$y_{i,t} = \sum_{\tau \in [-5,5]; \tau \neq -1} \beta_\tau \cdot (\mathbf{1}[t - T(s) = \tau] \cdot \mathbf{1}(grp = HighRisk) + \lambda_{s,T(s),r,m,grp} + \gamma_{t,T(s)} + \varepsilon_{i,t},$$

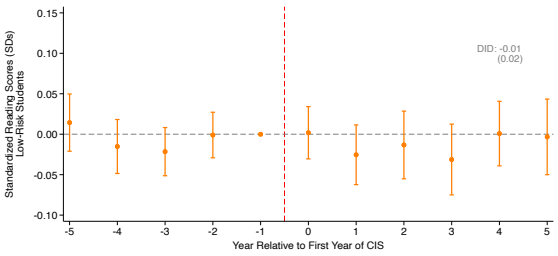
where grp indexes whether student i is a high- or low-risk student. We include the same set of year fixed effects (γ) as in the matched control design and interact the school fixed effects (λ) with an indicator for being high-risk to account for the large level difference in outcomes between high- and low-risk students (see Table II). Panel A restricts to treated schools, comparing outcomes for high- versus low-risk students as the CIS program begins. Panel B repeats the same analysis in matched control schools, where no program is implemented. The coefficients reflect differences in test score trajectories between high- and low-risk students before and after program rollout, netting out school-by-year shocks. Standard errors are clustered at the school-by-birth-cohort level. We report DID coefficients from a similarly modified version of Equation 2.

FIGURE A.6: Additional Test Score Results

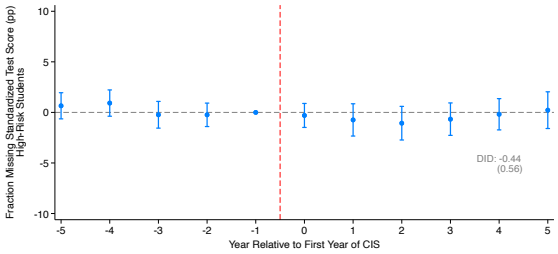
A: Effect of CIS on Std. Reading Scores
High-Risk Students



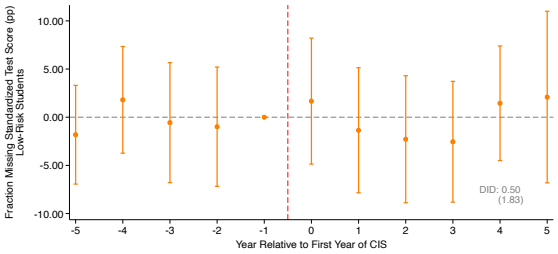
B: Effect of CIS on Std. Reading Scores
Low-Risk Students



C: Effect of CIS on Missing Std. Math
High-Risk Students



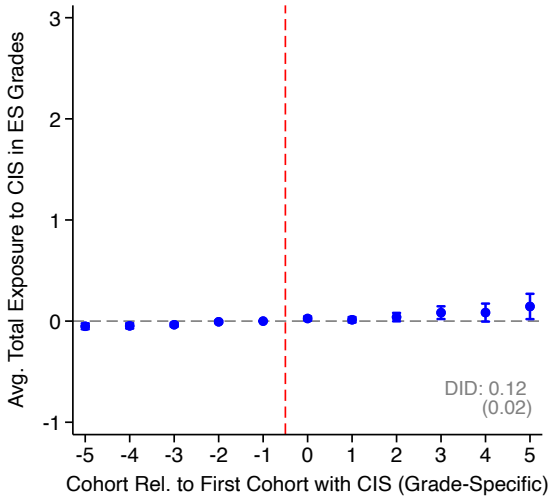
D: Effect of CIS on Missing Std. Math
Low-Risk Students



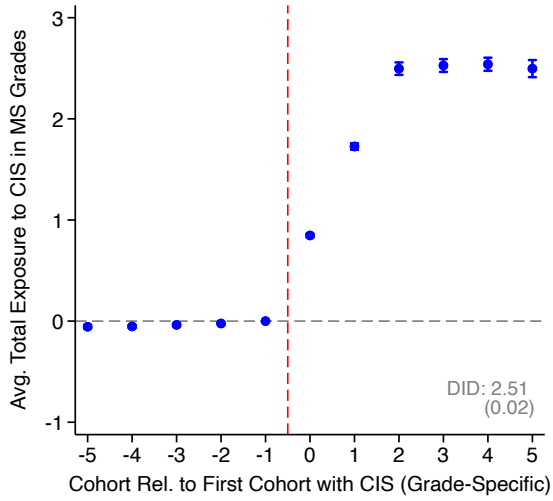
Notes: This figure presents additional results on the effect of CIS on student test score outcomes. Panels A and B show event studies analogous to those in Figure V, but for standardized reading scores instead of standardized math scores. In Panels C and D, the specification is the same, but the outcome is an indicator variable equal to 1 if a student is missing a standardized math score in year t. See notes for Figure V for additional details.

FIGURE A.7: Detailed First Stage by School Type

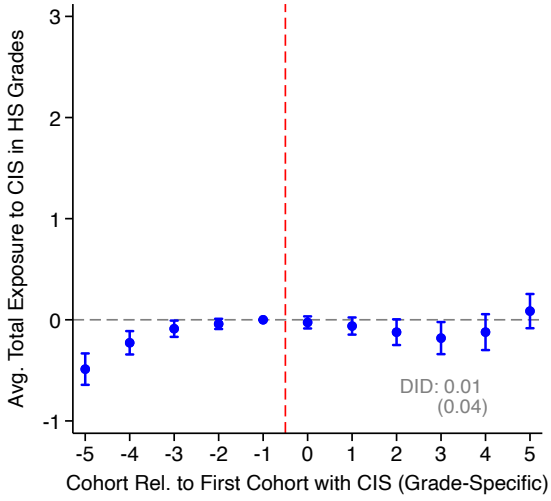
A: Total Exposure to CIS in Elementary School



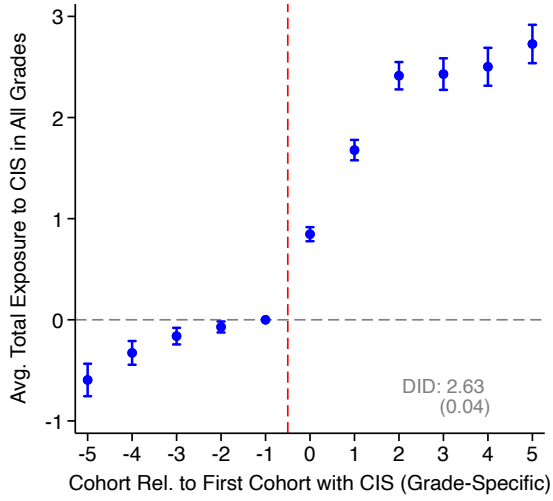
B: Total Exposure to CIS in Middle School



C: Total Exposure to CIS in High School

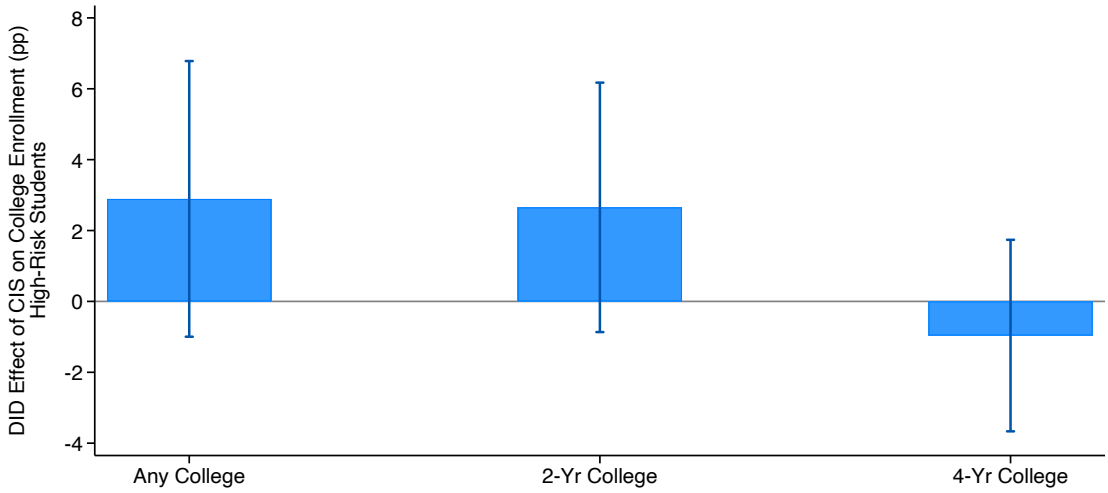


D: Total Exposure to CIS in All Grades



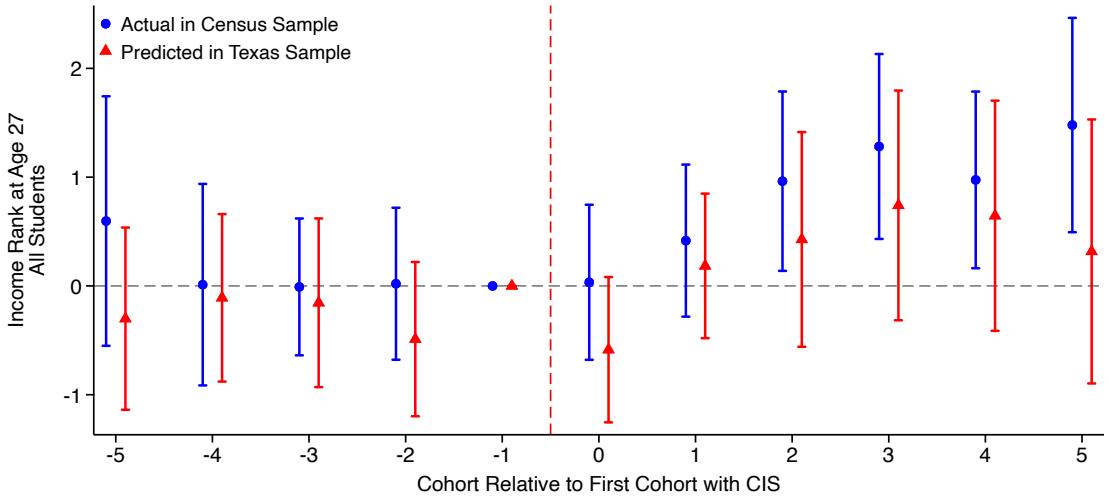
Notes: This figure presents event study estimates of CIS exposure by school level. In each panel, we estimate Equation 3, where the dependent variable is cumulative years of exposure to CIS during elementary school (Panel A), middle school (Panel B), high school (Panel C), and across all grades (Panel D). All specifications use the matched control design and leverage variation from the rollout of CIS to middle schools, as described in Section IV. These estimates assess whether variation in middle school exposure is correlated with exposure at other grade levels. Standard errors are clustered at the school-by-birth-cohort level.

FIGURE A.8: Effect of CIS on College Enrollment by College Type
High-Risk Students



Notes: This figure presents reduced form, DID coefficients from three separate regressions showing the effect of CIS on college attendance for high-risk students. The middle bar is the same as that shown in Figure VIII Panel *C* and restricts to two-year college attendance, see notes for additional details. The bar on the left captures college attendance at all levels and the bar on the right captures enrollment only at 4-year colleges. In each case, we define enrollment as ever enrolling between the ages of 18-22. Standard errors are clustered at the school-by-birth-cohort level.

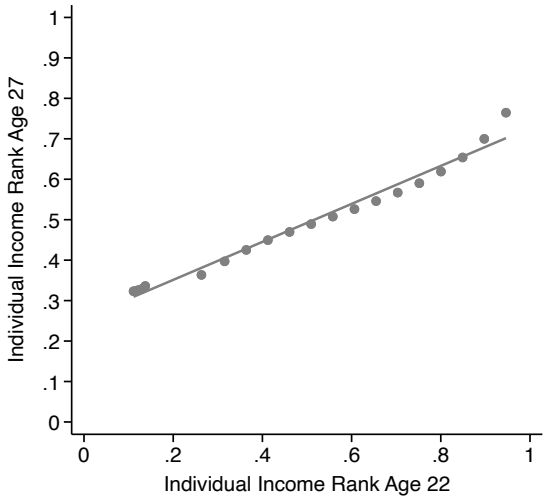
FIGURE A.9: Sample Comparison: Effect of CIS Income Rank at Age 27
All Students



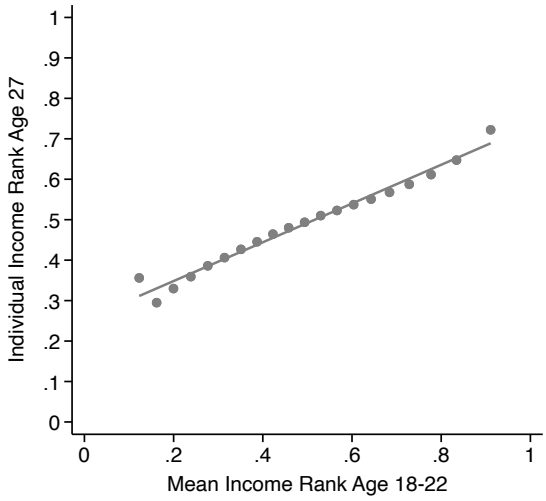
Notes: This figure compares the effect of CIS on adult earnings rank as measured in the national Census-IRS data to the effect on predicted earnings in the Texas sample. The Census series (circles) replicates the estimates from Figure IX; see accompanying notes for additional details. The Texas series (triangles) presents coefficients from Equation 4, where the dependent variable is predicted earnings rank at age 27. These predictions are based on observed earnings and college attendance between ages 18 and 22. See Appendix Figure A.10 for further details on the earnings prediction model. Standard errors in the Census sample are clustered at the high school catchment level. Standard errors in the Texas sample are clustered at the school level.

FIGURE A.10: Predicting Individual Income Rank at Age 27

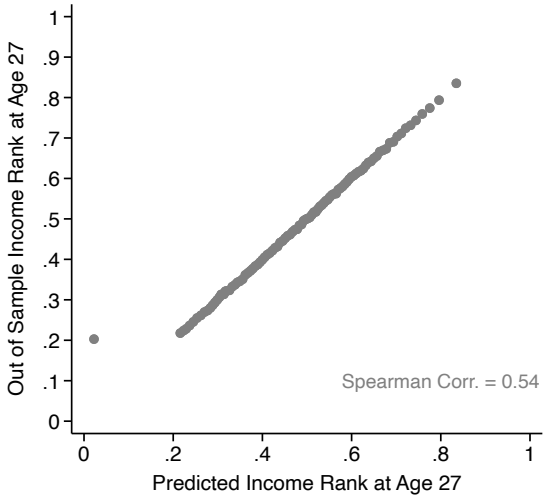
A: Relationship with Age 22 Rank



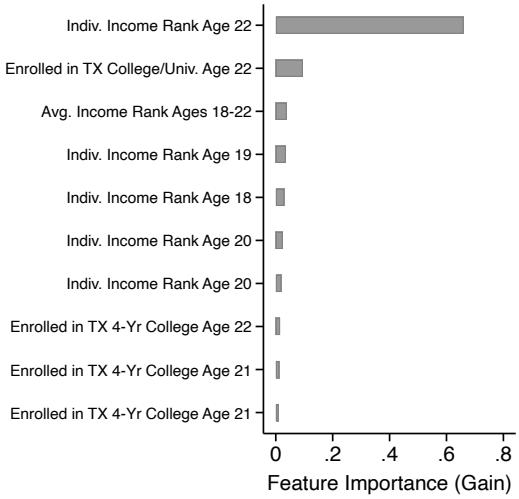
B: Relationship with Mean Rank Ages 18-22



C: Out-of-Sample Accuracy of Age 27 Income Rank

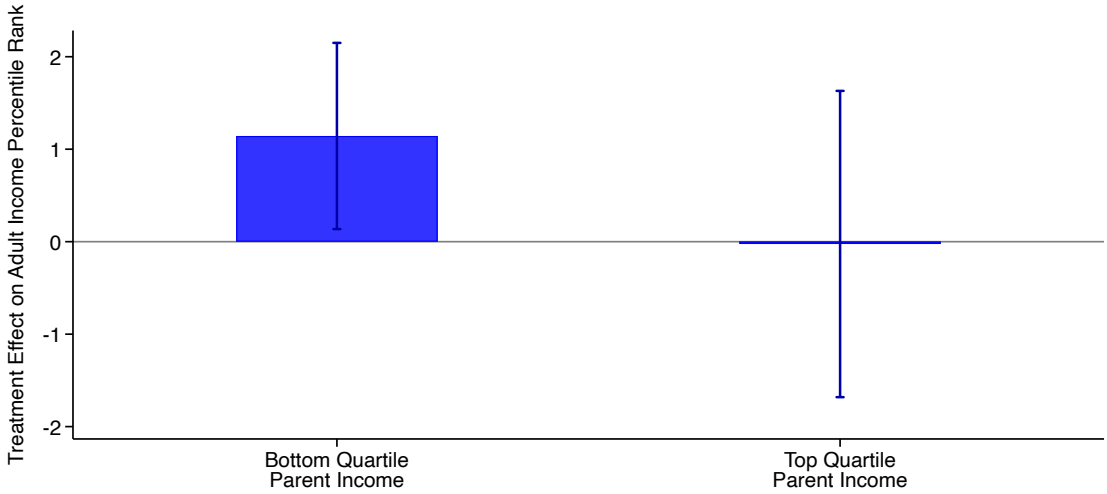


D: Feature Importance – Model Gain



Notes: This figure summarizes the empirical prediction of household income rank at age 27 using earnings and college attendance data from ages 18 to 22. All plots are based on a 20% hold-out test set not used in model training, using statewide data students born between 1978-1994. Panel A shows a binned scatterplot of the relationship between income rank at age 22 and income rank at age 27. Panel B presents the analogous relationship using mean income rank across ages 18 to 22. Panel C shows a binned scatterplot of observed income ranks against predicted values from a gradient boosting model trained on earlier outcomes. Predictors include income rank at each age (18–22), average rank, change in rank over time, and indicators for 2-year and 4-year college attendance at each age. Panel D reports the top 10 predictors by average model gain.

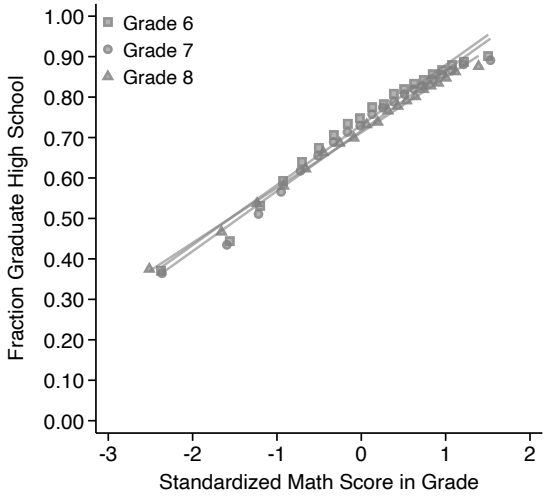
FIGURE A.11: Effect of CIS Household Income Rank at Age 27
Heterogeneity by Parent Income Quartile



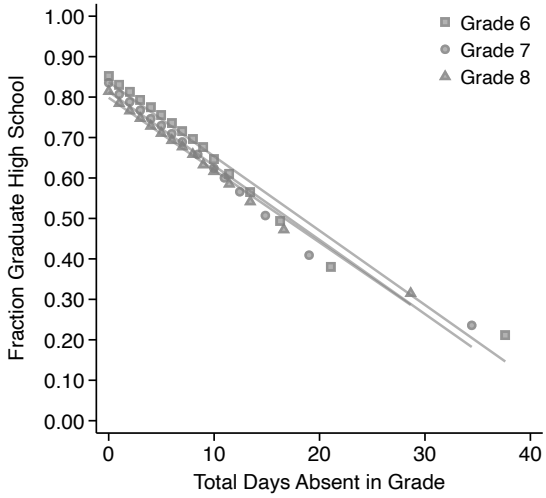
Notes: This figure displays the reduced form DID effect of CIS on adult earnings rank as measured in the national Census-IRS data, separately for children with parents in the bottom and top quartiles of the parent income distribution. Standard errors are clustered at the high school catchment zone boundary. Census Bureau Disclosure Review Board no. CBDRB-FY23-CES014-028 (Goldman, Gracie and Porter, 2023).

FIGURE A.12: Predicting High School Graduation

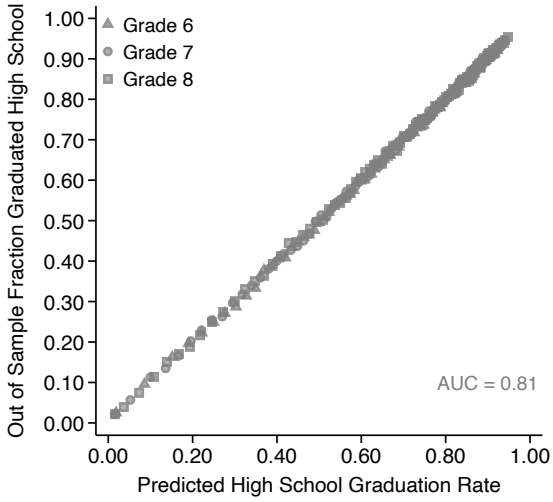
A: Relationship with Test Scores



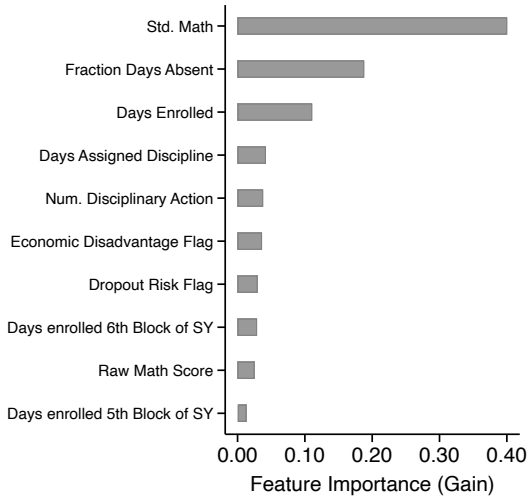
B: Relationship with Attendance



C: Out-of-Sample Accuracy of HS Graduation Prediction



D: Feature Importance – Model Gain



Notes: This figure summarizes the empirical prediction of high school graduation using academic and behavioral outcomes observed in middle school. All plots are based on a 20% hold-out test set not used in model training, using statewide data on middle school students from 1998–2004. Panel A shows a binned scatterplot of the relationship between standardized math scores and high school graduation, separately for grades 6–8. Panel B presents the analogous relationship using days absent in each grade. Panel C shows a binned scatterplot of observed graduation rates against predicted graduation probabilities from grade-specific gradient boosting models. These models are trained on student characteristics and outcomes measured in grade g, including demographics (race, sex, FRPL, dropout-risk flag, economic disadvantage flag), standardized math scores, attendance (at six-week intervals and overall), and disciplinary records (in-/out-of-school suspensions, days assigned, number of infractions). Panel D reports the top 10 predictors by average model gain across the grade-specific models.