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)	Fraction of Case-Managed Students-Years
	(1)	(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

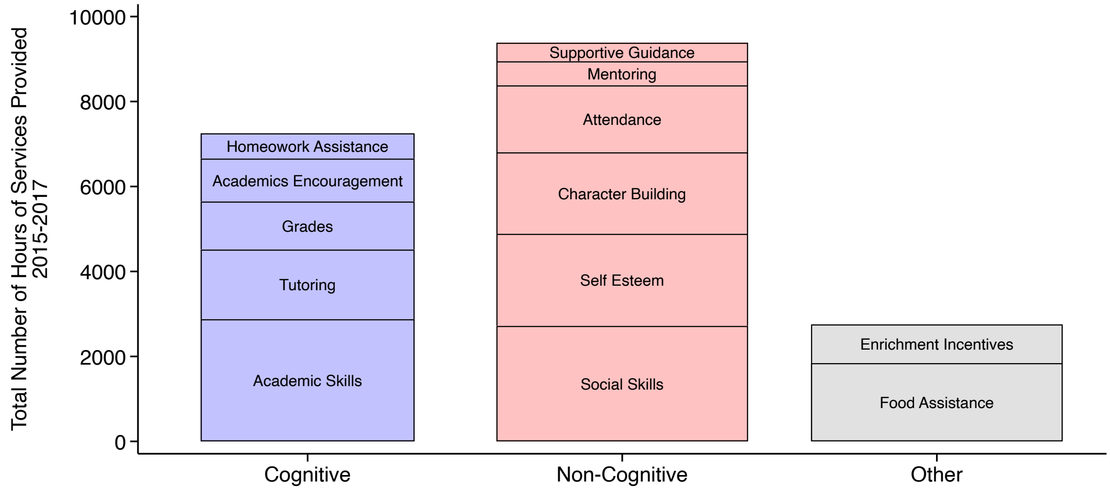
Notes: This table provides information on the detailed needs of students who are case managed by CIS in middle schools from 2014-2019, using the CIS microdata described in Section III B. For each need category, we provide the fraction of student-year observations who are flagged as having this need (not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

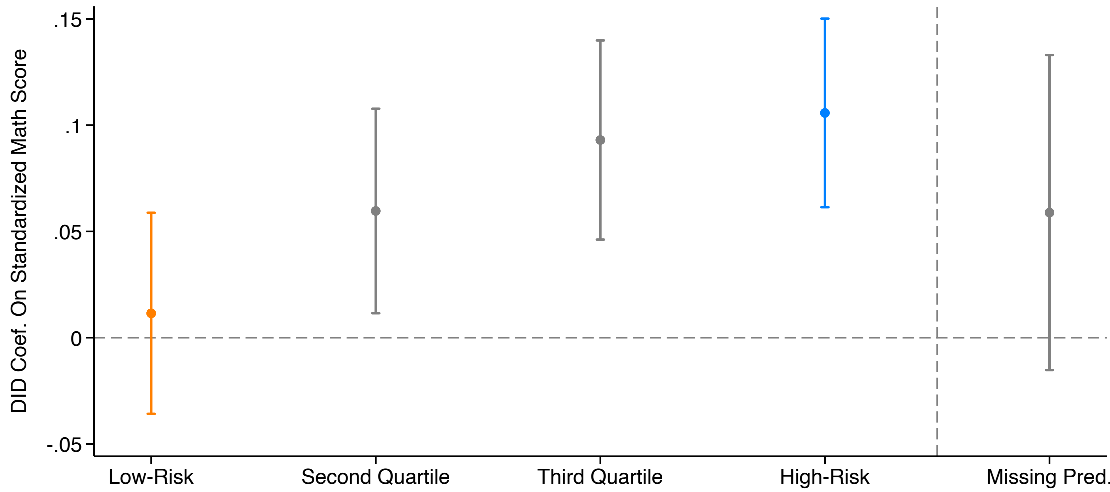
Notes: This table presents results from the same regression as those in Table III, but with standard errors clustered at the school level. See table notes for Table III for additional details.

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



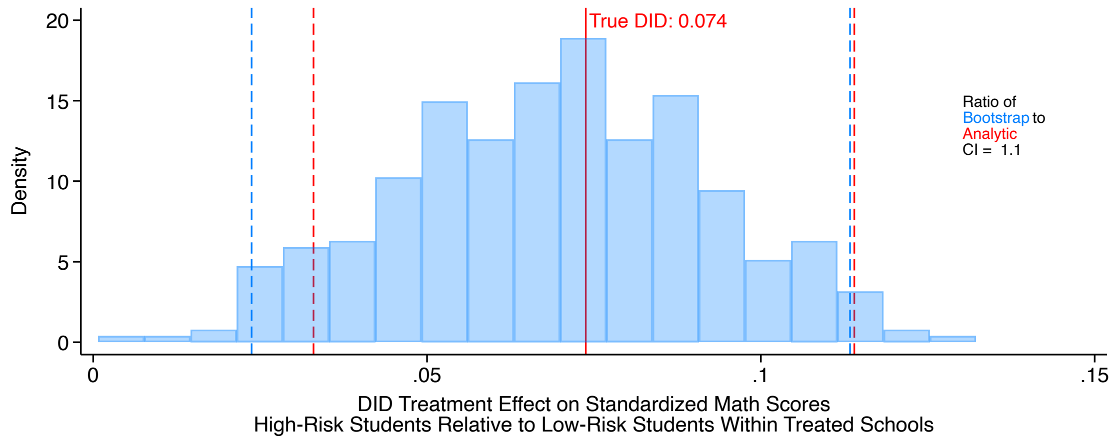
Notes: This figure summarizes the most common services provided by CIS navigators, measured in total time. Using the CIS microdata described in Section III/B, we identify the top 10 most frequently provided services in each year from 2015 to 2017. We classify these services into three broad categories 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.

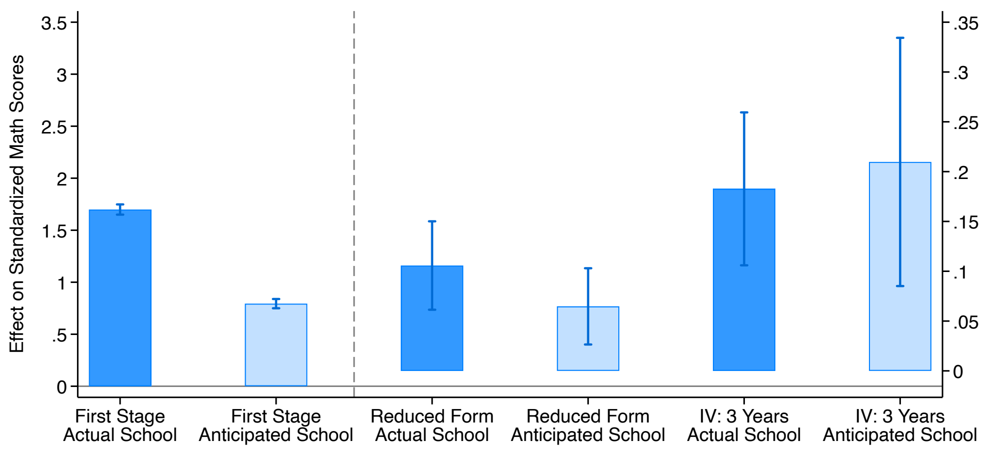
FIGURE A.3: Clustered Bootstrap: Within-School Effects on Test Scores
High-Risk Students



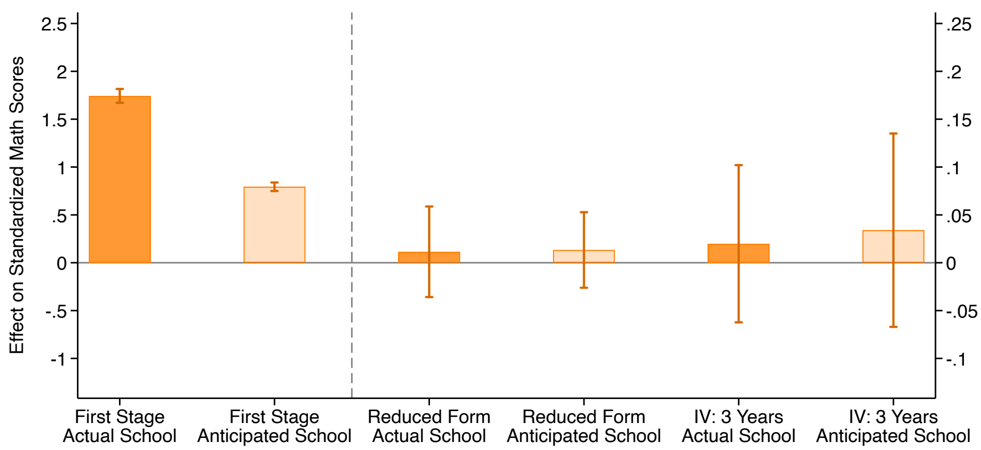
Notes: This figure shows the distribution of treatment effects from a clustered bootstrap procedure. For computational reasons, we do not re-estimate the full prediction model from Figure III or the corresponding treatment effects from Figure V. Instead, we take a simplified approach based on the within-school 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



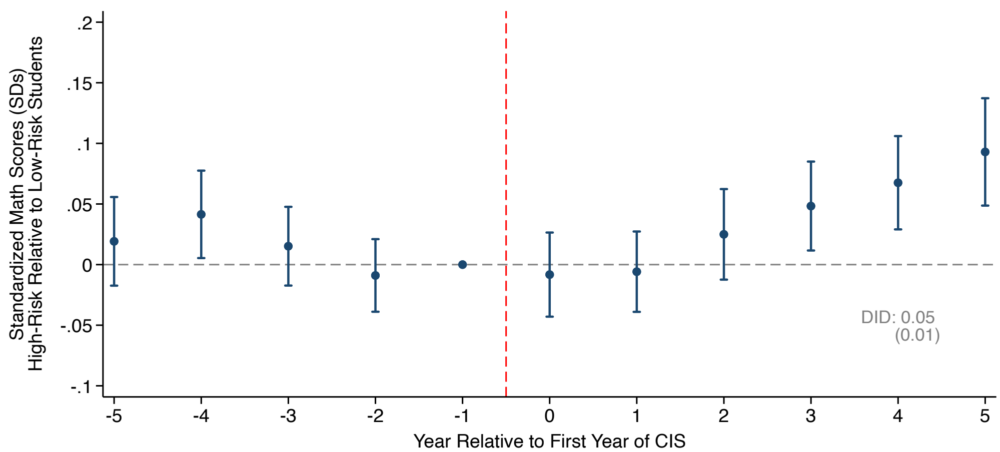
B: Low-Risk Students



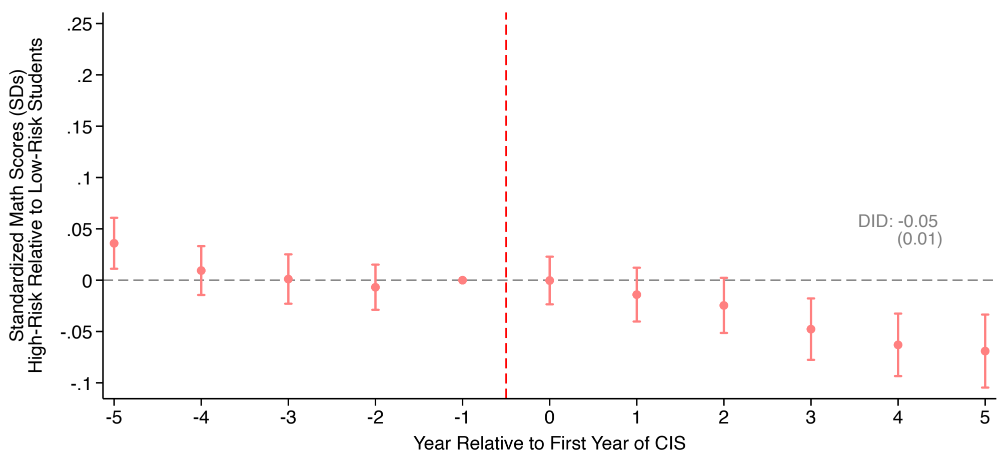
Notes: This figure compares the first stage, reduced form, and IV effects of CIS on standardized test scores using students' actual versus anticipated middle school assignments. In Panel A, we restrict to high-risk students. The first stage and reduced form estimates using actual school assignment replicate 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



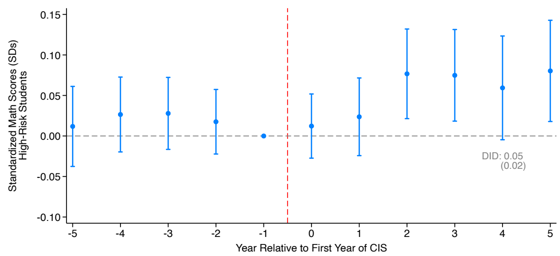
Notes: This figure presents event study coefficients estimating the relative effect of CIS on standardized test scores for high-risk students compared to their low-risk peers within the same school. To do this, we estimate a specification similar to Equation 1, but we modify it to compare outcomes for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

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

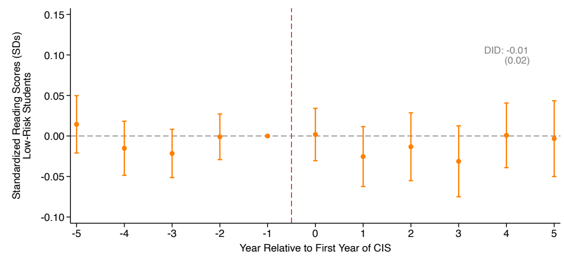
where grp indexes whether student i is a high- or low-risk student. We include the same set of year fixed effects (γ) as in the matched control design and interact the school fixed effects (λ) with an indicator for being high-risk to account for the large level difference in outcomes between high- and 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

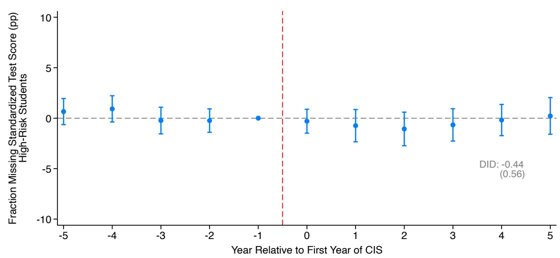
A: Effect of CIS on Std. Reading Scores
High-Risk Students



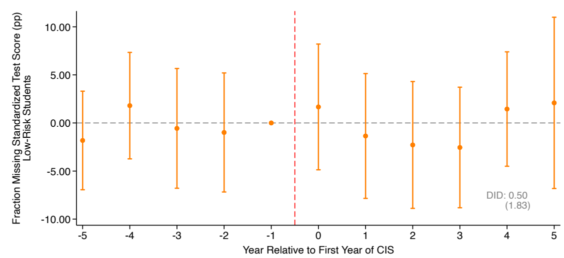
B: Effect of CIS on Std. Reading Scores
Low-Risk Students



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



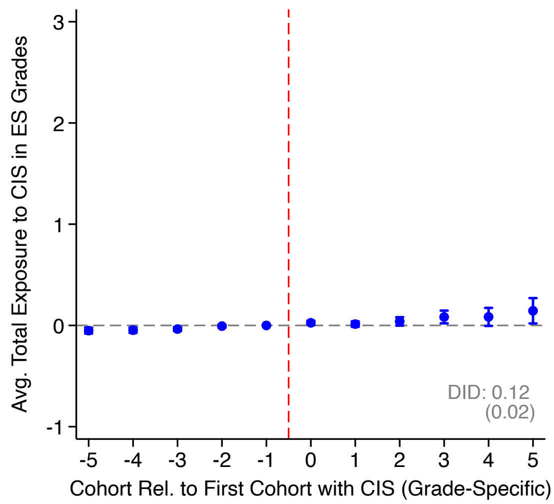
D: Effect of CIS on Missing Std. Math
Low-Risk Students



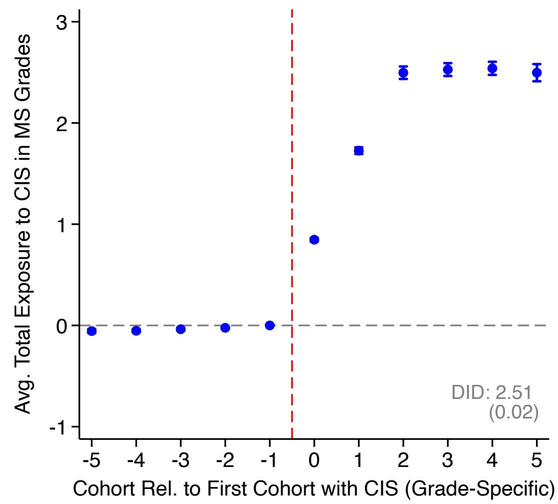
Notes: This figure presents additional results on the effect of CIS on student test score outcomes. Panels A and B show event studies analogous to those in Figure V, but for standardized reading scores instead of standardized math scores. In Panels C and D, the specification is the same, but the outcome 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

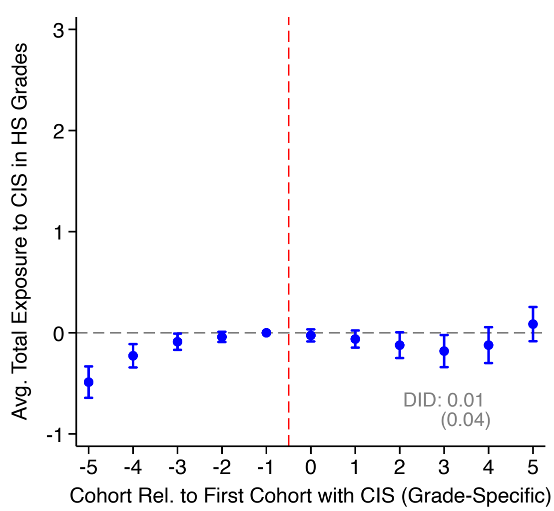
A: Total Exposure to CIS in Elementary School



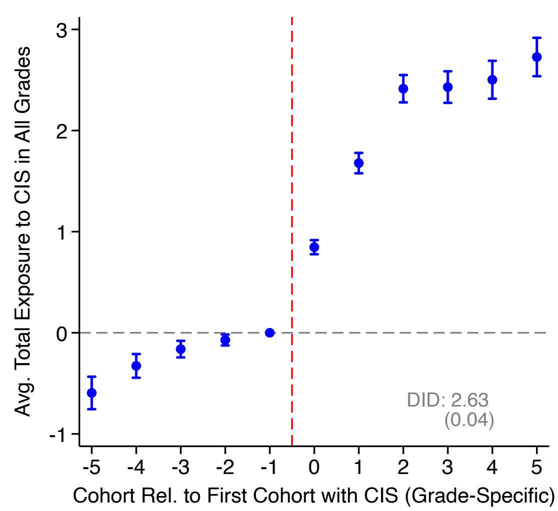
B: Total Exposure to CIS in Middle School



C: Total Exposure to CIS in High School

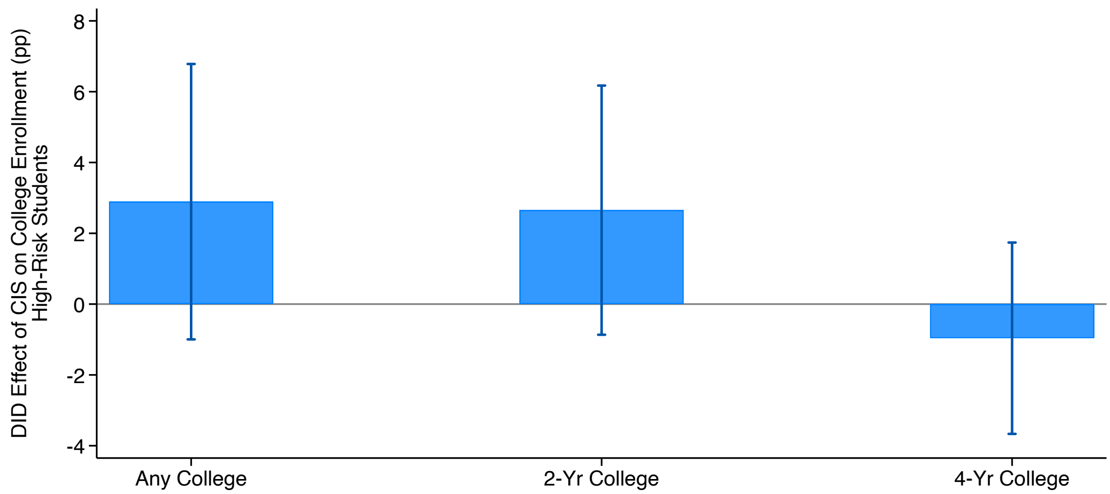


D: Total Exposure to CIS in All Grades



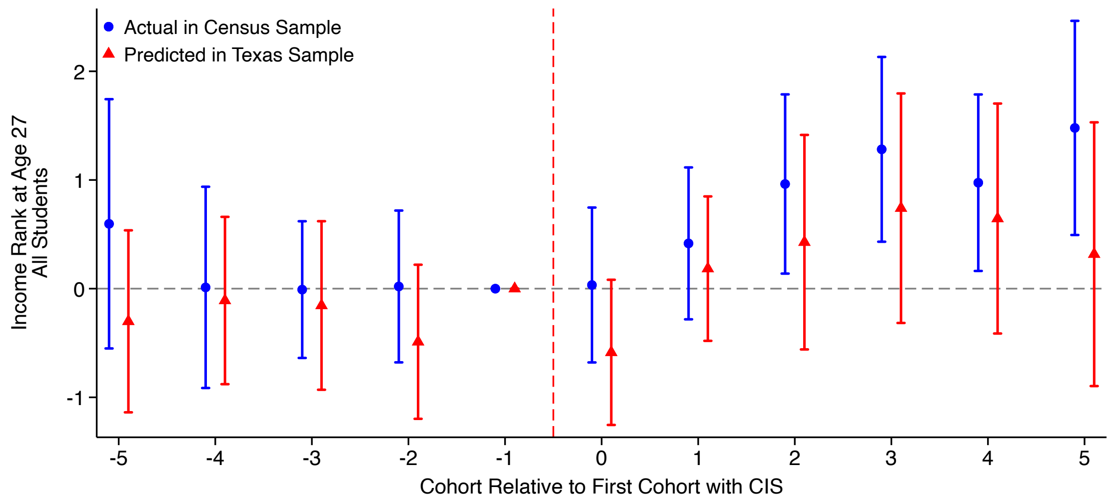
Notes: This figure presents event study estimates of CIS exposure by school level. In each panel, we estimate Equation 3, where the dependent variable is cumulative years of exposure to CIS during elementary school (Panel A), middle school (Panel B), high school (Panel C), and across all grades (Panel 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



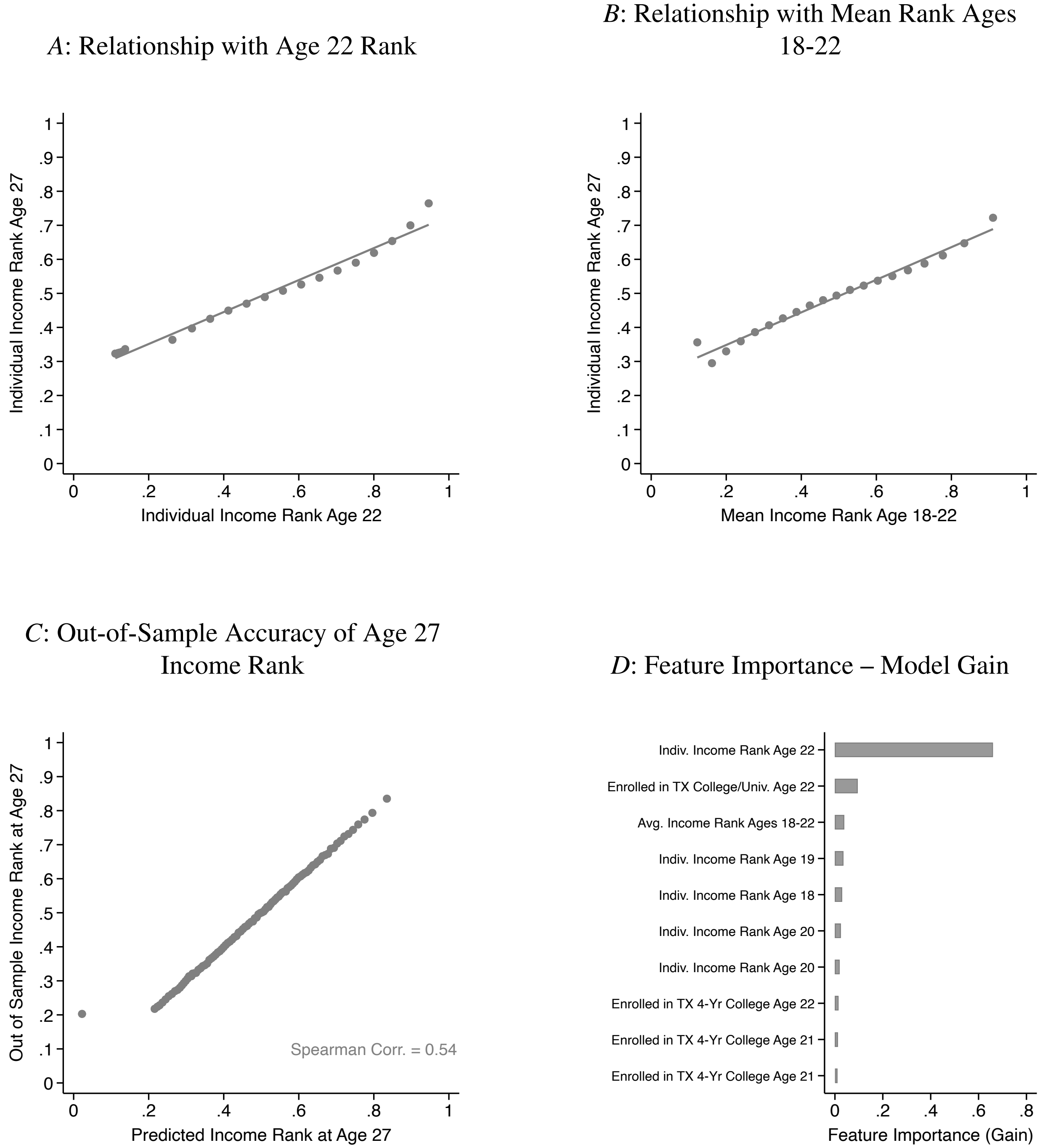
Notes: This figure presents reduced form, DID coefficients from three separate regressions showing the effect of CIS on college attendance for high-risk students. The middle bar is the same as that shown in Figure VIII Panel *C* and restricts to two-year college attendance, see notes for additional details. 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



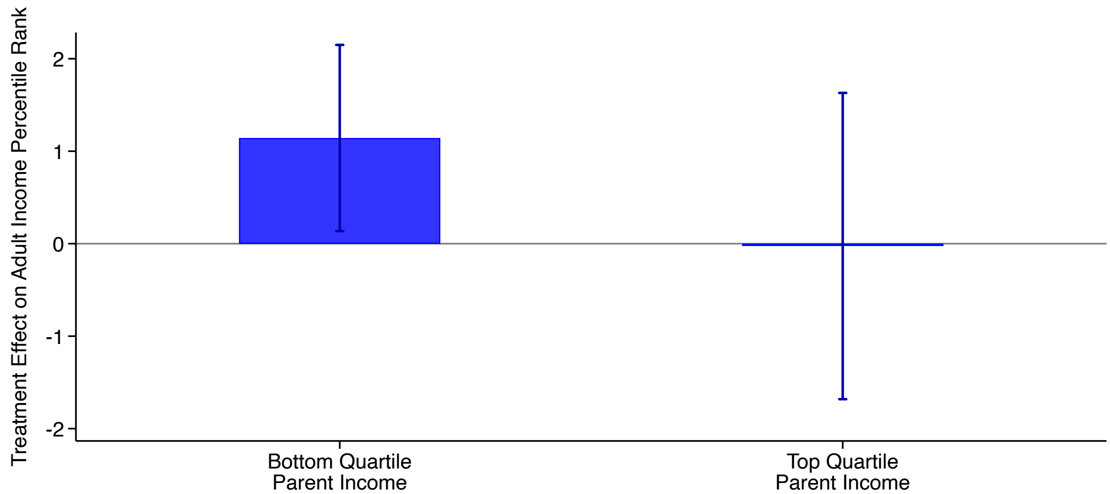
Notes: This figure compares the effect of CIS on adult earnings rank as measured in the national Census-IRS data to the effect on predicted earnings in the Texas sample. The Census series (circles) replicates the estimates from Figure IX; see accompanying notes for additional details. The Texas series (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



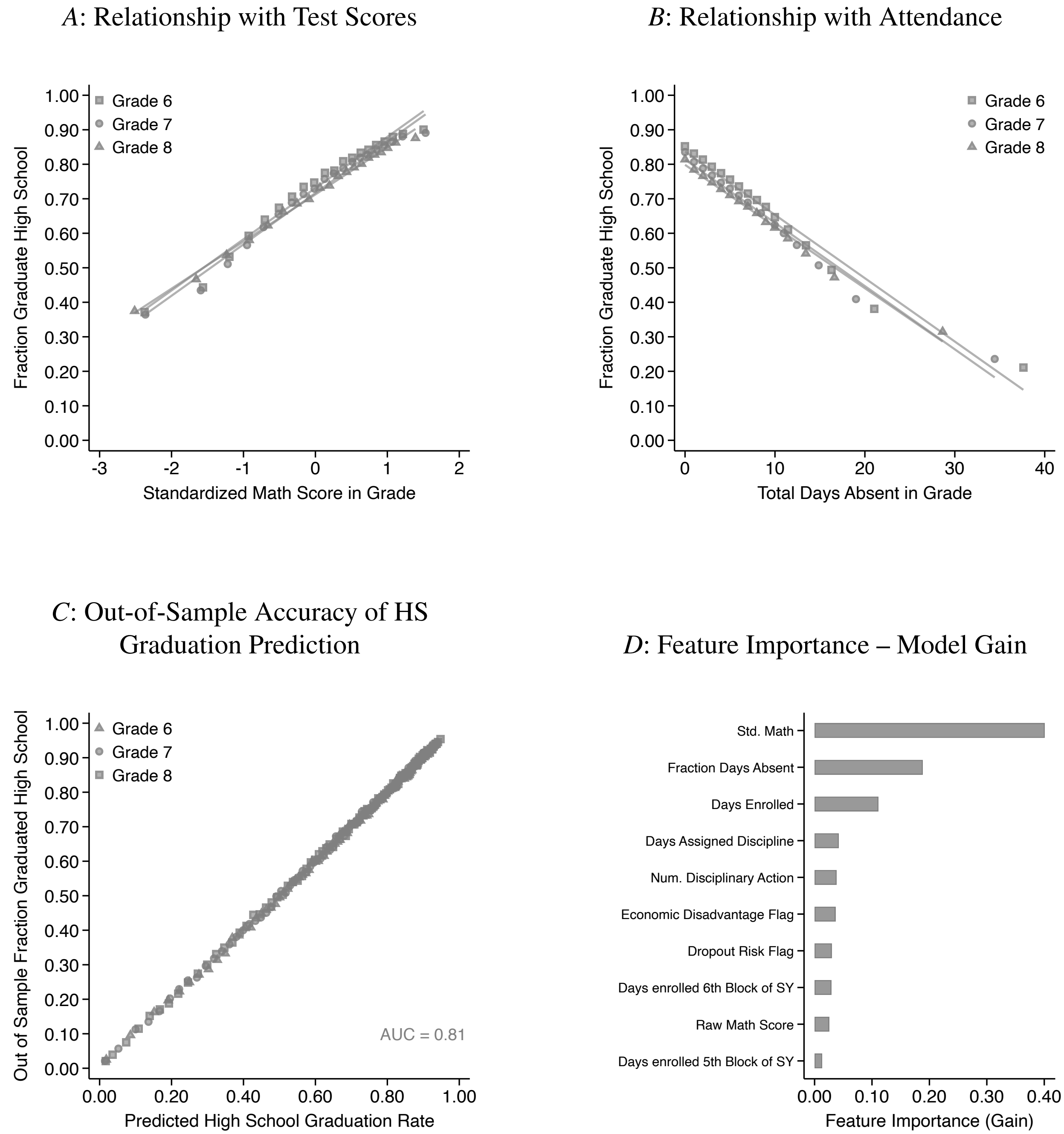
Notes: This figure summarizes the empirical prediction of household income rank at age 27 using earnings and college attendance data from ages 18 to 22. All plots are based on a 20% hold-out test set not used in model training, using statewide data students born between 1978-1994. Panel A shows a binned 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



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.