Saga Paper Exhibits

March 2021

Table 1. Baseline Test Score Comparison of Study Samples vs. All CPS Students

A. Study 1 Sample: Pre-randomization Math Scores (School Year 2012-2013)

Sample	N	Mean	25th	50th	75th
All CPS Students^	409258	44.26	19	43	68
All CPS 9th/10th Grade Boys	31064	44.76	19	41	71
All Study School 9th/10th Grade Students	4406	35.4	14.75	33	54
All Study School 9th/10th Grade Boys	2434	35.22	14	29	54
All Randomized Students*	2633	37.12	14	33	59
All Randomized Students Active in Study Schools	2103	36.32	14	33	54
All Participating Treatment Students	526	32.2	12	29	45

B. Study 2 Sample: Pre-randomization Math Scores (School Year 2013-2014)

Sample	N	Mean	25th	50th	75th
All CPS Students^	406672	47.16	22	46	72
All CPS 9th/10th Grade Students	61824	48.73	23	49	76
All Study School 9th/10th Grade Students	5068	34.89	12	29	54
All Randomized Students*	2645	33.56	13	30	51
All Randomized Students Active in Study Schools	1823	31.15	12	28	49
All Participating Treatment Students	571	30.05	10	26	47

C. Study 2 Sample by Gender: Pre-randomization Math Scores (School Year 2013-2014)

Sample	N	Mean	25th	50th	75th
All CPS Students^	406672	47.16	22	46	72
All CPS 9th/10th Grade Students	61824	48.73	23	49	76
All Study School 9th/10th Grade Girls	2384	34.05	13	29	53.75
All Randomized Girls*	799	30.26	11	28	47
All Randomized Girls Active in Study Schools	629	28.81	8	26	44
All Participating Treatment Students: Girls	144	28.11	8.5	26	42
All Study School 9th/10th Grade Boys	2684	35.68	12	29	54
All Randomized Boys*	1847	34.99	15	32	53
All Randomized Boys Active in Study Schools	1194	32.4	12	28	49
All Participating Treatment Students: Boys	427	30.72	11	26	49

In study 1, 1307 students were randomized into the treatment group and 1326 to the control group. In study 2, 1546 students (459 female students and 1088 male students)†were randomized into the treatment group; 1125 students were randomized into the control group (349 female students and 776 male students). In study 1, 15 students randomized to the control group received one or more session of Saga. In study 2, 88 students randomized to control (15 female students and 73 male students) participated in one or more session of Saga programming.

Due to student mobility and other factors *N*=65 students were randomized into study 2 twice, with 39 (60%) retaining their original treatment assignment across randomizations and 26 being randomized into the other treatment condition. Of these 26 students, 12 (46%) participated in one or more session of Saga programming.

∗Randomization was done using summer rosters of partner schools based on who they expected to enroll; a subset of these students showed up in these schools in the Fall of the study year. We refer to the students who do not show up as 'No Shows', and we include them in the main analysis. We also present our main analyses excluding these students in the appendices (appendix tables 12 and 13).

^The N for 'All CPS Students' refers to all active CPS students in the network. We use baseline tests from the prior year for all CPS students in grades 3-11 for whom tests were available for this row.

†For randomization, we use gender and grade at the time of randomization (neither of which are necessarily time invariant) for blocking.

Table 2: Baseline Characteristics for Saga Study 1 and Study 2 Cohorts by Randomized Treatment Group

	Study 1 Control Mean, N=1326	Treat/Control Contrast	Robust Standard Error	Study 2 Control Mean, N=1145	Treat/Control Contrast	Robust Standard Error
Demographics						
Black	0.461	-0.003	0.011	0.643	-0.006	0.012
Latinx	0.487	0.011	0.013	0.326	-0.001	0.013
ELL	0.120	0.015	0.013	0.103	0.008	0.011
Female	0.002	0	0.001	0.307	0.012*	0.007
Age	14.807	-0.036	0.025	14.430	0.007	0.022
Free and Reduced Lunch recipient	0.871	0.002	0.013	0.902	-0.019	0.012
Learning Disability	0.167	0.012	0.015	0.159	-0.002	0.014
Schooling						
In Grade 9 at Study Start	0.572	-0.02*	0.011	0.914	-0.008	0.006
In Grade 10 at Study Start	0.416	0.02*	0.011	0.075	0.007	0.005
GPA in Baseline Year	2.109	0.009	0.036	2.386	-0.043	0.031
Pct A's Earned in Baseline Year	18.774	0.331	0.801	24.744	-0.94	0.707
Pct B's Earned in Baseline Year	22.382	-0.49	0.655	23.922	0.085	0.573
Pct C's Earned in Baseline Year	26.781	-0.097	0.708	28.825	-0.294	0.619
Pct D's Earned in Baseline Year	16.486	0.85	0.646	11.625	0.738	0.470
Pct F's Earned in Baseline Year	15.578	-0.594	0.916	10.884	0.411	0.665
Math Test Performance in Baseline Year	0.000	-0.04	0.042	0.000	-0.026	0.041
Crime						
Ever Arrested Baseline	0.187	-0.009	0.015	0.148	0.008	0.014
Baseline Total Arrests	0.517	-0.028	0.061	0.422	-0.065	0.059
Number of baseline arrests for:						
Violent offenses	0.131	0.007	0.021	0.117	-0.014	0.019
Property offenses	0.090	-0.016	0.017	0.067	-0.006	0.015
Drug offenses	0.070	-0.004	0.015	0.046	-0.009	0.011
Other offenses	0.226	-0.015	0.032	0.190	-0.036	0.031

F-test for treatment-control comparison for all baseline characteristics: **Study 1**: p=0.832; **Study 2**: p=0.893; **Pooled Study 1 and 2**: p=0.828

Baseline treatment/control comparisons, controlling for randomization blocks. Heteroskedasticity-robust standard errors calculated.
Randomization blocks are defined at the school-grade level in Study 1 (boys only) and school-grade-gender level with the inclusion of girls in study 2.
Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the table above.
Percent of study participants female: Study 1: 0.1; Study 2: 30.8; Pooled Study 1 and 2: 15.5. Study 1 eligibility was limited to males.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.

Table 3: Estimated Effects on Academic and Behavioral Outcomes in Study 1, Year 1

	N	Control Mean	Intent to Treat Estimate	Robust Standard Error	Effect of Treatment on Treated (TOT)	Robust Standard Error	Control Complier Mean	False Discovery Rate Q-Value
Mathematics								
SY 2014 Math Test Scaled Score (Z)	1852	-0.000	0.082**	(0.036)	0.16**	(0.069)	-0.091	0.032
SY 2014 Math GPA	2215	1.760	0.274***	(0.04)	0.561***	(0.079)	1.628	0.001
SY 2014 Math Courses Failed (percent)	2215	0.191	-0.043***	(0.013)	-0.087***	(0.026)	0.179	0.003
ISR Math Test 2014 Z-Score	617	-0.000	0.119**	(0.058)	0.188*	(0.094)	-0.088	0.058
ISR Math Test 2014 Scaled Score	617	31.110	1.344**	(0.566)	2.183**	(0.927)	29.882	0.032
Other Academics								
SY 2014 Read Test Scaled Score (Z)	1851	0.000	0.004	(0.04)	0.011	(0.076)	-0.115	0.882
SY 2014 Non-Math GPA	2244	1.739	0.081**	(0.033)	0.201***	(0.063)	1.724	0.005
SY 2014 Non-Math Core Courses Failed (percent)	2244	0.210	-0.027**	(0.011)	-0.056**	(0.022)	0.220	0.021
Behavior								
SY 2014 Disciplinary Incidents	2494	1.513	0.092	(0.104)	0.208	(0.235)	1.485	0.399
SY 2014 Days Absent	2494	24.242	0.7	(0.831)	1.58	(1.874)	22.790	0.399
SY 2014 Out-of-School Suspensions	2494	1.515	0.199	(0.153)	0.457	(0.345)	1.522	0.399
Crime								
SY 2014 Arrests for Violent Crimes	2633	0.086	-0.015	(0.015)	-0.037	(0.035)	0.103	0.708
SY 2014 Arrests for Property Crimes	2633	0.061	-0.01	(0.01)	-0.024	(0.025)	0.055	0.708
SY 2014 Arrests for Drug Crimes	2633	0.057	0.02	(0.014)	0.048	(0.033)	-0.004	0.708
SY 2014 Arrests for Other Crimes	2633	0.178	-0.002	(0.022)	-0.005	(0.053)	0.121	0.931
SY 2014 Ever Arrested	2633	0.176	-0.007	(0.013)	-0.018	(0.03)	0.147	0.834
SY 2014 Total Arrests	2633	0.382	-0.008	(0.037)	-0.018	(0.087)	0.275	0.931

Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included. Randomization blocks are defined at the school-grade level.
Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
"Ever Arrested" indicates whether the student had any arrests during the period.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives. Each panel in the above table is a 'family', so for example, 'Mathematics' represents one outcome family. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995); the results are qualitatively similar if we use the method from Benjamini, Krieger and Yekutieli (2006).

Table 4: Estimated Effects on Academic and Behavioral Outcomes in Study 2, Year 1

	N	Control Mean	ITT Estimate	Clustered Standard Error	Effect of Treatment on Treated (TOT)	Clustered Standard Error	Control Complier Mean	False Discovery Rate Q-Value
Mathematics								
SY 2015 Math Test Scaled Score (Z)	1865	0.008	0.126***	(0.036)	0.369***	(0.108)	-0.144	0.002
SY 2015 Math GPA	2062	1.859	0.15***	(0.043)	0.42***	(0.124)	1.787	0.002
SY 2015 Math Courses Failed (percent)	2062	0.149	-0.029**	(0.013)	-0.082**	(0.037)	0.187	0.026
Other Academics								
SY 2015 Read Test Scaled Score (Z)	1865	0.007	-0.007	(0.04)	-0.007	(0.119)	-0.057	0.954
SY 2015 Non-Math GPA	2110	1.936	0.065*	(0.034)	0.182*	(0.1)	1.934	0.206
SY 2015 Non-Math Core Courses Failed (percent)	2110	0.138	-0.011	(0.01)	-0.031	(0.028)	0.165	0.424
Behavior								
SY 2015 Disciplinary Incidents	2474	1.554	-0.002	(0.138)	-0.035	(0.434)	1.759	0.936
SY 2015 Days Absent	2474	22.777	0.756	(0.822)	2.194	(2.59)	23.126	0.691
SY 2015 Out-of-School Suspensions	2474	0.732	0.07	(0.091)	0.211	(0.286)	0.564	0.691
Crime								
SY 2015 Violent Crime Arrests	2710	0.104	-0.013	(0.016)	-0.04	(0.052)	0.138	0.541
SY 2015 Property Crime Arrests	2710	0.072	-0.026	(0.016)	-0.09	(0.055)	0.137	0.209
SY 2015 Drug Crime Arrests	2710	0.051	0	(0.012)	0.001	(0.04)	0.035	0.983
SY 2015 Other Crime Arrests	2710	0.225	-0.052*	(0.028)	-0.168*	(0.092)	0.341	0.203
SY 2015 Ever Arrested	2710	0.164	-0.019	(0.013)	-0.06	(0.042)	0.195	0.237
SY 2015 Total Arrests	2710	0.452	-0.091**	(0.045)	-0.297**	(0.15)	0.651	0.203

Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above.
To account for individuals having multiple observations, standard errors are clustered on individuals.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included. Randomization blocks are defined at the school-grade-gender level.
Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
"Ever Arrested" indicates whether the student had any arrests during the period.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives. Each panel in the above table is a 'family', so for example, 'Mathematics' represents one outcome family. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995); the results are qualitatively similar if we use the method from Benjamini, Krieger and Yekutieli (2006).

Table 5: Estimated 1 Year Treatment Effects on Academic and Behavioral Outcomes: Pooling Study 1 and 2

	N	Control Mean	ITT Estimate	Clustered Standard Error	Effect of Treatment on Treated (TOT)	Clustered Standard Error	Control Complier Mean	False Discovery Rate Q-Value
Mathematics								
Math Test Scaled Score (Z)	3717	0.004	0.113***	(0.026)	0.264***	(0.061)	-0.125	0.001
Math GPA	4277	1.803	0.219***	(0.029)	0.518***	(0.07)	1.674	0.001
Math Courses Failed (percent)	4277	0.173	-0.037***	(0.009)	-0.087***	(0.022)	0.185	0.001
Other Academics								
Read Test Scaled Score (Z)	3716	0.003	0	(0.028)	0.007	(0.067)	-0.093	0.920
Non-Math GPA	4354	1.825	0.077***	(0.024)	0.185***	(0.058)	1.825	0.005
Non-Math Core Courses Failed (percent)	4354	0.178	-0.02***	(0.007)	-0.048***	(0.018)	0.198	0.012
Behavior								
Disciplinary Incidents	4968	1.532	0.047	(0.086)	0.108	(0.23)	1.613	0.638
Days Absent	4968	23.581	0.767	(0.582)	1.817	(1.55)	22.908	0.362
Out-of-School Suspensions	4968	1.161	0.136	(0.089)	0.347	(0.237)	1.071	0.362
Crime								
Arrests for Violent Crimes	5343	0.094	-0.013	(0.011)	-0.036	(0.031)	0.118	0.292
Arrests for Property Crimes	5343	0.066	-0.018*	(0.01)	-0.051*	(0.028)	0.091	0.292
Arrests for Drug Crimes	5343	0.054	0.009	(0.009)	0.027	(0.025)	0.014	0.292
Arrests for Other Crimes	5343	0.200	-0.026	(0.018)	-0.07	(0.051)	0.216	0.292
Ever Arrested	5343	0.171	-0.012	(0.009)	-0.031	(0.025)	0.165	0.292
Total Arrests	5343	0.414	-0.048	(0.029)	-0.13	(0.083)	0.440	0.292

Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above.
To account for individuals having multiple observations, standard errors are clustered on individuals.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included. Randomization blocks are defined at the school-grade-gender level.
Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
"Ever Arrested" indicates whether the student had any arrests during the period.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives. Each panel in the above table is a 'family', so for example, 'Mathematics' represents one outcome family. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995)

Table 6: Estimated Effects of Year 2 Treatment on Academic and Behavioral Outcomes in Study 1

	N	Control Mean	ITT Estimate	Standard Error	Effect of Treatment in Year 2 on Treated (TOT)	Standard Error	False Discovery Rate Q-Value	Effect of Treatment in Year 1 and/or Year 2 on Treated (TOT)	Standard Error	False Discovery Rate Q-Value
Mathematics										
SY 2015 Math Test Scaled Score (Z)	1640	-0.000	0.163***	(0.038)	0.839***	(0.198)	0.001	0.304***	(0.07)	0.001
SY 2015 Math GPA	1841	1.870	0.14***	(0.049)	0.685***	(0.234)	0.009	0.263***	(0.089)	0.010
SY 2015 Math Courses Failed (percent)	1841	0.168	-0.021	(0.015)	-0.103	(0.07)	0.182	-0.04	(0.027)	0.183
ISR Math Test 2015 Z-Score	878	0.000	0.067	(0.052)	0.303	(0.232)	0.198	0.112	(0.086)	0.200
ISR Math Test 2015 Scaled Score	878	33.178	0.984*	(0.563)	4.459*	(2.496)	0.127	1.651*	(0.918)	0.133
Other Academics										
SY 2015 Read Test Scaled Score (Z)	1640	0.000	-0.055	(0.044)	-0.286	(0.227)	0.577	-0.104	(0.082)	0.537
SY 2015 Non-Math GPA	1895	1.860	0.035	(0.041)	0.107	(0.19)	0.577	0.041	(0.073)	0.537
SY 2015 Non-Math Core Courses Failed (percent)	1895	0.180	-0.007	(0.012)	-0.037	(0.059)	0.577	-0.014	(0.023)	0.537
Behavior										
SY 2015 Disciplinary Incidents	2251	1.126	-0.036	(0.105)	-0.206	(0.579)	0.760	-0.076	(0.215)	0.761
SY 2015 Days Absent	2251	23.141	-0.24	(0.899)	-1.523	(4.956)	0.760	-0.566	(1.839)	0.761
SY 2015 Out-of-School Suspensions	2251	0.716	0.049	(0.095)	0.265	(0.521)	0.760	0.098	(0.194)	0.761
Crime										
SY 2015 Arrests for Violent Crimes	2633	0.089	0	(0.015)	0	(0.097)	0.998	0	(0.035)	0.998
SY 2015 Arrests for Property Crimes	2633	0.057	0.002	(0.012)	0.012	(0.08)	0.998	0.004	(0.029)	0.998
SY 2015 Arrests for Drug Crimes	2633	0.084	0.025	(0.02)	0.16	(0.127)	0.279	0.058	(0.046)	0.282
SY 2015 Arrests for Other Crimes	2633	0.187	0.036	(0.028)	0.237	(0.179)	0.279	0.086	(0.065)	0.282
SY 2015 Ever Arrested	2633	0.159	0.019	(0.013)	0.125	(0.083)	0.279	0.045	(0.03)	0.282
SY 2015 Total Arrests	2633	0.418	0.063	(0.047)	0.409	(0.301)	0.279	0.148	(0.109)	0.282

Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included. Randomization blocks are defined at the school-grade level.
Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
"Ever Arrested" indicates whether the student had any arrests during the period.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives. Each panel in the above table is a 'family', so for example, 'Mathematics' represents one outcome family. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995)

Table 7: Impacts on Self-Reported Risky Behavior and Crime Victimization by Study 1 subjects: end of Second Program Year

	N	Control Mean	Intention to Treat	FDR Q-value
Risky Behavior				
During your life, how many days have you had at least one drink of alcohol? (Z)	888	-0.000	-0.207***	0.019
			(0.061)	
During the past 30 days, on how many days did you have at least one drink of alcohol? (Z)	890	-0.000	-0.187***	0.019
			(0.063)	
During your life, how many times have you used marijuana? (Z)	884	-0.000	-0.089	0.380
			(0.066)	
During the past 30 days, how many times did you use marijuana? (Z)	886	-0.000	-0.074	0.449
			(0.065)	
During your life, how many times have you tried any other sort of illegal drug/inhalant/prescription drug? (Z)	889	-0.000	-0.132**	0.152
			(0.066)	
Do any of your brothers, sisters, cousins, or friends belong to a gang? (Dummy)	887	0.318	-0.008	0.834
			(0.032)	
Do you belong to a gang? (Dummy)	889	0.079	-0.017	0.526
			(0.017)	
Have you ever sold marijuana or any other drug to your friends? (Dummy)	888	0.133	-0.029	0.380
			(0.022)	
Have you ever sold marijuana or any other drug to people you didn't know? (Dummy)	888	0.105	-0.025	0.380
			(0.019)	
During the past 3 months with how many people did you have sexual intercourse? (Z)	557	0.000	-0.202	0.380
			(0.178)	
How many times have you gotten someone pregnant? (Z)	558	0.000	0.074	0.526
			(0.078)	
In the past year, how many times did you get in a physical fight in which you were so badly injured that you were treated by a doctor or a nurse? (Z)	895	0.000	0.015	0.834
			(0.063)	
In the past year, how often did you hurt someone badly enough in a physical fight that he or she needed to be treated by a doctor or nurse? (Z)	895	0.000	-0.143**	0.121
			(0.063)	
During the past 30 days, on how many days did you carry a weapon – such as a gun, knife, or club – to school? (Z)	892	-0.000	-0.049	0.526
			(0.064)	
In the past year, how often did you paint graffiti or signs on someone else's property or in a public place? (Z)	895	0.000	-0.054	0.526
			(0.061)	
In the past year, how often did you deliberately damage property that didn't belong to you? (Z)	896	0.000	-0.056	0.526
			(0.066)	
In the past year, how often did you take something from a store without paying for it? (Z)	894	-0.000	0.014	0.834
			(0.068)	
In the past year, how often did you drive a car without owner's permission? (Z)	895	0.000	0.027	0.827
			(0.071)	
In the past year, how often did you break into someone's home in order to steal? (Z)	893	-0.000	-0.069	0.380
			(0.051)	
Crime Victimization				
In the past year, how often did someone pulle a gun/knife on you? (Z)	894	-0.000	0.006	0.922
			(0.069)	
In the past year, how often did you get into a physical fight? (Z)	894	0.000	-0.056	0.848
			(0.066)	
In the past year, how often did you get jumped? (Z)	896	0.000	0.031	0.848
			(0.067)	
In the past year, how often did you get beaten up and something was stolen from you? (Z)	894	-0.000	0.043	0.848
			(0.083)	

Note: Data are from survey designed by DETAIL and given to a randomly selected subsample of youth, proportional to overall treatment and control group size randomized into Saga during 2015. All items are coded so the desired effect direction is positive.
Baseline covariates and randomization block fixed effects included in all models (see text). Heteroskedasticity-robust standard errors in parentheses. * = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives. Each panel in the above table is a 'family'. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995)

Table 8: 11th Grade Outcomes: Saga Studies 1 and 2, pooled

Outcome	N	Control Mean	Intent-to-Treat Effect	ITT Standard Error	Treatment-on-treated Effect	TOT Standard Error	Control Complier Mean	FDR Q-value
Held Back by 11th Grade	3333	0.102	0.002	(0.01)	0.004	(0.022)	0.088	0.872
Mathematics								
11th Grade Math Test (z-score)	2973	0.006	0.096***	(0.029)	0.222***	(0.068)	-0.137	0.003
11th Grade GPA: Math	3086	1.980	0.075**	(0.037)	0.179**	(0.086)	1.843	0.037
Non-math Academics								
11th Grade GPA: All Non-Math Courses	3339	1.944	0.022	(0.031)	0.057	(0.072)	1.794	0.782
11th Grade Reading Test (z-score)	2972	0.007	0.009	(0.031)	0.02	(0.073)	-0.123	0.782

Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above.
To account for individuals having multiple observations, standard errors are clustered on individuals.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included. Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives.
Each panel in the above table is a 'family', so for example, 'Mathematics' represents one outcome family. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995).

Table 9: Graduation outcomes: Saga Studies 1 and 2, pooled

Outcome	N	Control Mean	Intent-to-Treat Effect	ITT Standard Error	Treatment-on-treated Effect	TOT Standard Error	Control Complier Mean	FDR Q-value
Graduated On-Time	3750	0.749	0.007	(0.013)	0.019	(0.034)	0.772	0.876
Ever Graduated	3770	0.823	0.004	(0.012)	0.008	(0.032)	0.859	0.876
Leave code: Dropout	5663	0.054	0.002	(0.007)	0.006	(0.019)	0.042	0.876
Leave code: Corrections	5663	0.028	0.003	(0.006)	0.009	(0.017)	0.011	0.876
Leave code: Transfer	5663	0.207	0.002	(0.012)	0.005	(0.033)	0.143	0.876
Leave code: Unknown	5663	0.134	-0.013	(0.01)	-0.038	(0.027)	0.154	0.766
Leave code: No exit	5663	0.026	0.001	(0.004)	0.003	(0.012)	0.023	0.876
Leave code: Deceased	5663	0.009	-0.004	(0.003)	-0.012	(0.009)	0.016	0.766

Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above.
To account for individuals having multiple observations, standard errors are clustered on individuals.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives.
We treat all outcomes above as one 'family'. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995).

Table 10: Study 1 Sample: Estimated Effects on Outcomes from ISR Survey - End of First Program Year

	N	Control Mean	Intention to Treat	FDR Q-value
Adult Supports				
Number of adults to talk to (No Change)	622	4.297	0.057	0.878
			(0.338)	
Number of adults who care (No Change)	623	7.384	0.355	0.878
			(0.594)	
Would talk to adults at school (Dummy)	623	0.375	-0.012	0.878
			(0.043)	
Grit				
Agree: Setbacks don't discourage me (Z)	623	-0.000	0.022	0.947
			(0.088)	
Agree: I am a hard worker (Z)	624	0.000	0.084	0.891
			(0.086)	
Disagree: I have difficulty maintaining focus (Z)	623	0.000	-0.092	0.891
			(0.082)	
Agree: I am diligent (Z)	624	-0.000	0.023	0.947
			(0.089)	
Agree: I finish what I begin (Z)	624	0.000	-0.051	0.947
			(0.087)	
Agree: I can continue until everything is perfect (Z)	624	-0.000	-0.006	0.947
			(0.088)	
Conscientiousness				
Agree: I am always prepared (Z)	624	0.000	0.118	0.507
			(0.089)	
Agree: I continue until everything is perfect (Z)	624	-0.000	-0.006	0.947
			(0.088)	
Agree: I leave a mess in my room (Z)	624	0.000	-0.011	0.947
			(0.087)	
Locus of Control				
Agree: I have control over direction of life (Z)	621	-0.000	0.052	0.564
			(0.088)	
Disagree: Every time I try to get ahead, something or somebody stops me (Z)	624	0.000	0.058	0.564
			(0.089)	
Disagree: Luck is more important than hard work (Z)	624	0.000	0.165**	0.125
			(0.087)	
Disagree: My plans never work out, planning makes me unhappy (Z)	622	-0.000	0.051	0.564
			(0.091)	
Agree: I can make plans work (Z)	623	-0.000	-0.219**	0.060
			(0.083)	
Social Networks				
Reports No Close Friends (Dummy)	623	0.025	-0.012	0.875
			(0.014)	
Friends think it is important to attend classes regularly (Z)	607	-0.000	-0.061	0.882
			(0.091)	
Friends think it is important to get good grades (Z)	607	-0.000	-0.035	0.967
			(0.082)	
Friends think it is important to study (Z)	607	0.000	-0.248***	0.049
			(0.091)	
Friends think it is important to continue education to college (Z)	607	-0.000	0.012	0.967
			(0.084)	
Have stopped hanging around with someone (Recoded Dummy)	623	0.505	0.066	0.438
			(0.045)	
Have started hanging around with someone (Recoded Dummy)	622	0.616	0.002	0.967
			(0.044)	

Note: Data are from survey designed by DETAIL and given to a randomly selected subsample of Study 1 youth, proportional to overall treatment and control group size randomized during 2015. All items are coded so the desired effect direction is positive. Baseline covariates and randomization block fixed effects included in all models (see text). Heteroskedasticity-robust standard errors in parentheses. * = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives. Each panel in the above table is a 'family'. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995)

Table 11: Study 1 Sample: Estimated Effects on Outcomes from ISR Survey - End of Second Program Year

	N	Control Mean	Intention to Treat	FDR Q-value
Adult Supports				
Number of adults to talk to (No Change)	893	4.089	0.143	0.567
			(0.252)	
Number of adults who care (No Change)	894	6.436	0.299	0.567
			(0.403)	
Grit				
Agree: Setbacks don't discourage me (Z)	896	0.000	-0.04	0.555
			(0.07)	
Agree: I am a hard worker (Z)	896	0.000	0.061	0.555
			(0.063)	
Disagree: I have difficulty maintaining focus (Z)	896	-0.000	-0.046	0.555
			(0.068)	
Agree: I am diligent (Z)	896	0.000	0.062	0.555
			(0.065)	
Conscientiousness				
I see myself as someone who does things carefully and completely.	896	-0.000	-0.023	0.886
			(0.066)	
I see myself as someone who can be somewhat careless.	895	-0.000	-0.023	0.886
			(0.069)	
I see myself as someone who is a reliable worker.	896	-0.000	0.029	0.886
			(0.068)	
I see myself as someone who tends to be disorganized.	896	0.000	-0.007	0.912
			(0.067)	
I see myself as someone who tends to be lazy.	896	0.000	-0.018	0.886
			(0.067)	
I see myself as someone who keeps working until things are done.	896	-0.000	0.054	0.886
			(0.063)	
I see myself as someone who does things efficiently (quickly and correctly).	896	-0.000	0.062	0.886
			(0.067)	
I see myself as someone who makes plans and sticks to them.	896	0.000	0.088	0.886
			(0.069)	
I see myself as someone who is easily distracted; has trouble paying attention.	896	-0.000	-0.055	0.886
			(0.067)	
Locus of Control				
Disagree: You can learn but not change intelligence (Z)	896	-0.000	0.117*	0.536
			(0.068)	
Disagree: Intelligence can't be changed much (Z)	896	-0.000	0.091	0.536
			(0.065)	
Disagree: You have a certain amount of intelligence, can't change (Z)	896	-0.000	0.069	0.564
			(0.067)	
Disagree: Moral character can't be changed (Z)	896	0.000	0.041	0.758
			(0.068)	
Disagree: Responsibility can't be changed (Z)	896	0.000	-0.016	0.813
			(0.069)	
Disagree: Moral traits can't be changed (Z)	896	0.000	0.067	0.564
			(0.069)	
Social Networks				
Reports No Close Friends (Dummy)	895	0.042	0.008	0.810
			(0.015)	
Friends think it is important to attend classes regularly (Z)	855	0.000	-0.018	0.810
			(0.071)	
Friends think it is important to get good grades (Z)	855	0.000	-0.057	0.810
			(0.067)	
Friends think it is important to study (Z)	855	0.000	0.025	0.810
			(0.068)	
Friends think it is important to continue education to college (Z)	853	-0.000	0.032	0.810
			(0.068)	
Have stopped hanging around with someone (Recoded Dummy)	895	0.593	-0.008	0.810
			(0.034)	
Have started hanging around with someone (Recoded Dummy)	896	0.644	-0.037	0.810
			(0.034)	

Note: Data are from survey designed by DETAIL and given to a randomly selected subsample of Study 1 youth, proportional to overall treatment and control group size randomized during 2015. All items are coded so the desired effect direction is positive. Baseline covariates and randomization block fixed effects included in all models (see text). Heteroskedasticity-robust standard errors in parentheses. * = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives. Each panel in the above table is a 'family'. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995)

Table 12: Impacts of Self-Reported Attitudes on Education

	N	Control Mean	Intention to Treat	FDR Q-value
Survey Wave 1: **School year 2013-2014**				
Disruptions by others get in the way of my learning (Z)	623	-0.000	-0.057	0.836
			(0.089)	
Weekly time spent on homework (Z)	623	0.000	-0.091	0.684
			(0.084)	
How much of your homework do you finish? (Z)	622	0.000	0.008	0.992
			(0.084)	
Expect to graduate high school and continue education (Dummy)	623	0.910	0.008	0.965
			(0.025)	
Want to graduate high school and continue education (Dummy)	623	0.923	0.002	0.992
			(0.024)	
How important are good grades to you? (Z)	623	0.000	0.061	0.836
			(0.083)	
Feeling safe at school (Z)	623	-0.000	0.041	0.938
			(0.094)	
Like math (Z)	623	-0.000	0.203**	0.085
			(0.086)	
Good grades in math (Z)	623	-0.000	0.289***	0.001
			(0.078)	
Like school in general (Z)	623	0.000	-0.13	0.577
			(0.086)	
Like reading (Z)	623	0.000	-0.102	0.684
			(0.082)	
Good grades in reading (Z)	623	0.000	-0.072	0.802
			(0.081)	
Survey Wave 2: **School year 2014-2015**				
Weekly time spent on homework (Z)	896	0.000	0.01	0.879
			(0.07)	
Expect to graduate high school and continue education (Dummy)	895	0.132	0.032	0.338
			(0.025)	
Want to graduate high school and continue education (Dummy)	896	0.630	0.037	0.338
			(0.032)	
How important are good grades to you? (Z)	896	-0.000	0.092	0.338
			(0.061)	
Feeling safe at school (Z)	896	-0.000	-0.019	0.869
			(0.07)	
Like math (Z)	896	-0.000	0.079	0.338
			(0.067)	
Good grades in math (Z)	896	-0.000	0.097	0.338
			(0.064)	
Like school in general (Z)	896	0.000	0.033	0.767
			(0.066)	
Like reading (Z)	896	-0.000	-0.083	0.338
			(0.067)	
Good grades in reading (Z)	895	0.000	-0.076	0.338
			(0.066)	

Note: Data are from survey designed by DETAIL and given to a randomly selected subsample of Study 1 youth, proportional to overall treatment and control group size randomized during 2015. All items are coded so the desired effect direction is positive. Baseline covariates and randomization block fixed effects included in all models (see text). Heteroskedasticity-robust standard errors in parentheses. * = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives. Each panel in the above table is a 'family'. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995)

Table 13: ITT Estimates Interacted with Whether School Above Median in CPS Disciplinary Incidents During School Year (Pooled)

Year 1 Outcome	N	Ctrl Mean	Intent-to-Treat Est.	ITT Standard Error	Dummy x Treatment Status	Interaction SE
Mathematics						
Math GPA	4331	1.786	0.246***	(0.043)	-0.05	(0.060)
Math Courses Failed (percent)	4331	0.179	-0.044***	(0.013)	0.011	(0.019)
Standardized Math Score	3741	0.001	0.13***	(0.038)	-0.031	(0.052)
Non-math academics						
Non-Math GPA	4409	1.809	0.062*	(0.035)	0.025	(0.049)
Non-math Core Courses Failed (percent)	4409	0.184	-0.019*	(0.010)	0.001	(0.015)
Standardized Reading Score	3740	0.001	-0.007	(0.041)	0.017	(0.056)
Behavior						
Disciplinary incidents	5035	1.549	-0.171**	(0.082)	0.379**	(0.166)
Days Absent	5035	23.932	-0.281	(0.719)	2.252*	(1.157)
Out-of-School Incs	5035	1.180	-0.09	(0.075)	0.421**	(0.174)
Crime						
Violent Crime Arrests	5410	0.095	0.005	(0.014)	-0.03	(0.021)
Property Crime Arrests	5410	0.073	-0.014	(0.013)	-0.014	(0.020)
Drug Arrests	5410	0.054	-0.013	(0.009)	0.043**	(0.017)
Other arrests	5410	0.207	-0.022	(0.022)	-0.008	(0.036)
Any Arrests (Dummy)	5410	0.176	-0.027**	(0.012)	0.029	(0.018)
All Arrests	5410	0.429	-0.045	(0.034)	-0.008	(0.058)

Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above. To account for individuals having multiple observations, standard errors are clustered on individuals. Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included. Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science). * = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.

Given that randomization block fixed-effects fully explain school fixed-effects, we cannot estimate the main effect for the 'above median' dummy variable, so we do not report this in the table. However, we can recover and report the interaction effect.

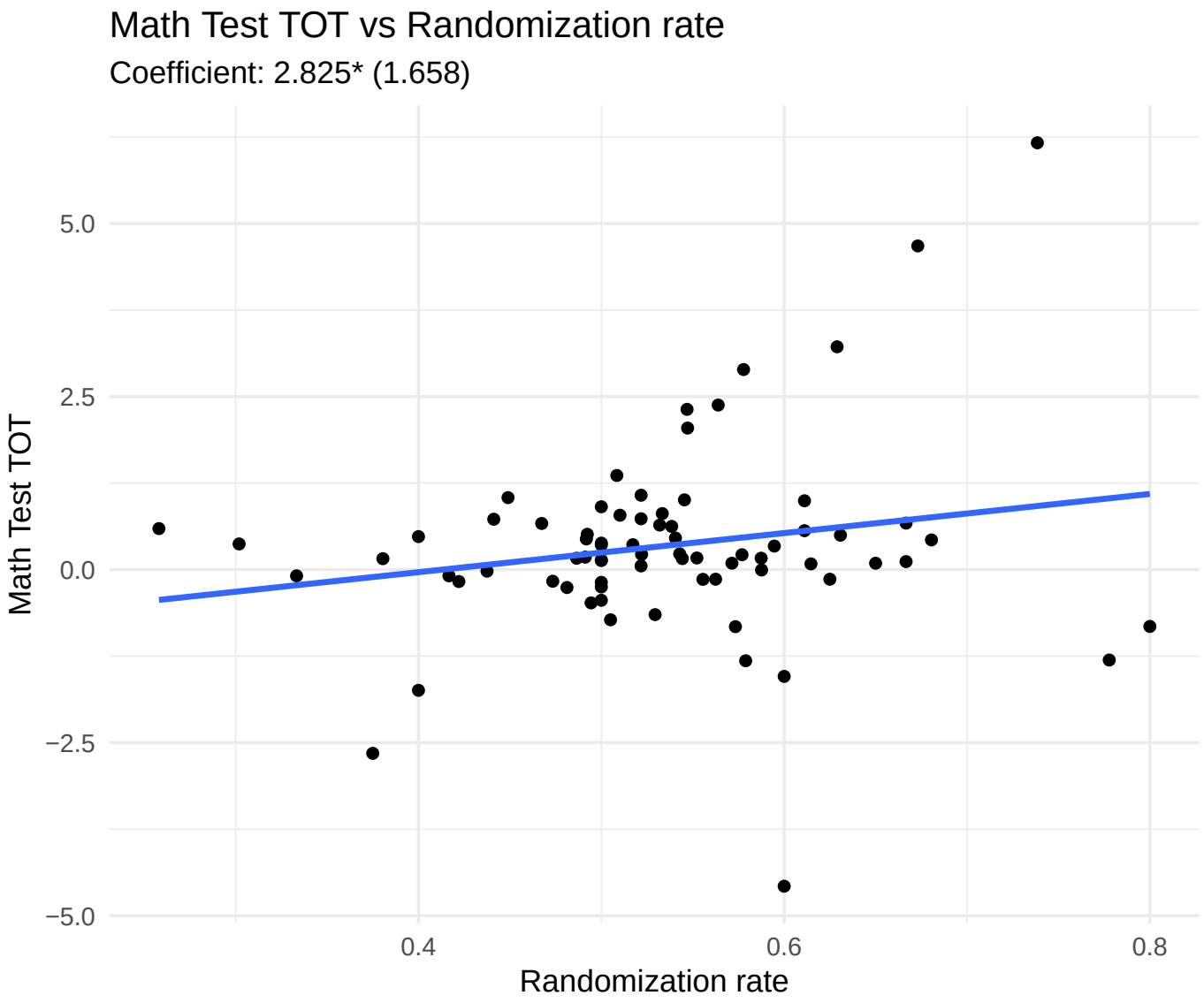

Figure 1: SUTVA Analysis: Block-level randomization rate plotted against TOT effect on math test score

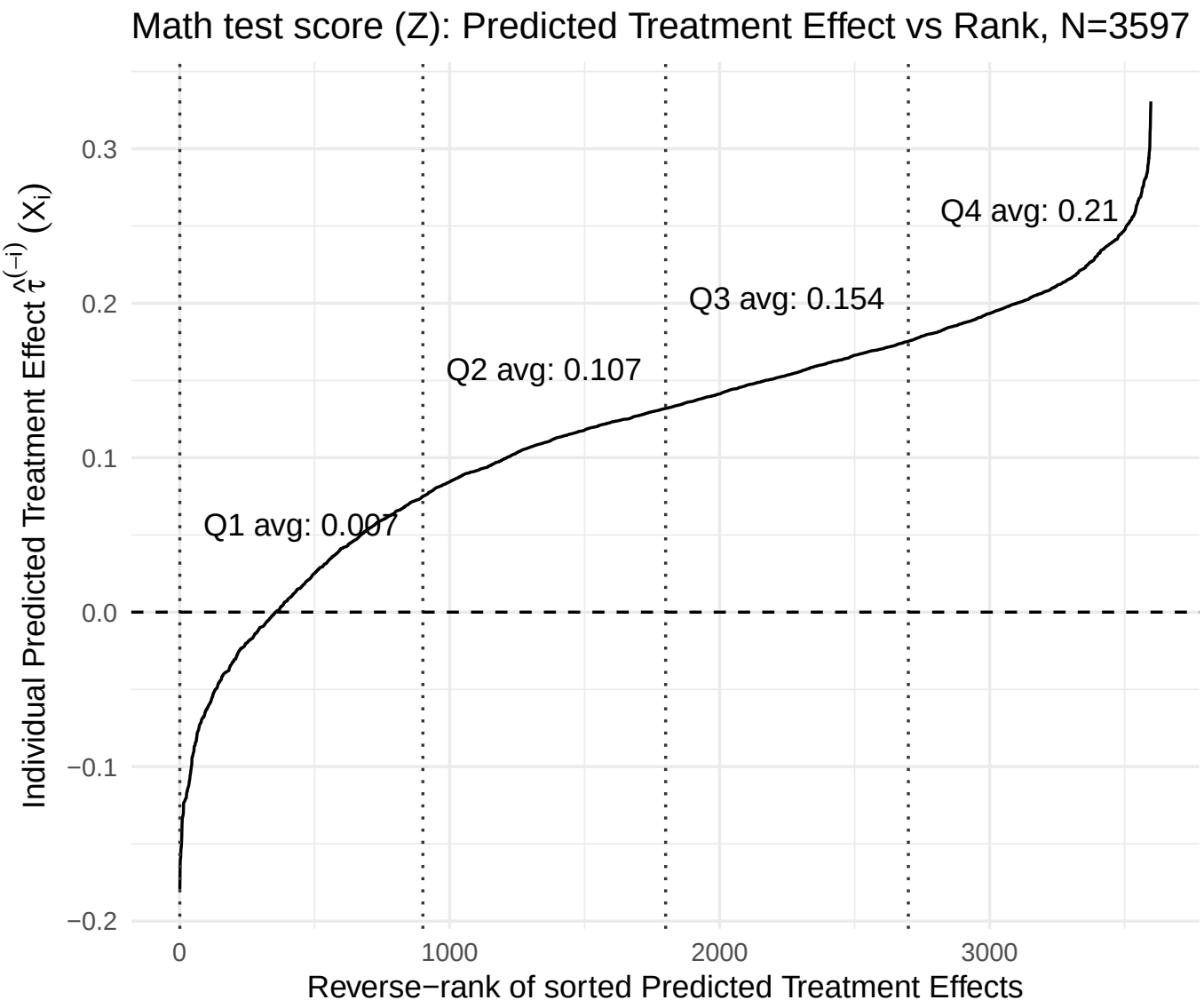

Figure 2: Random Forests for Treatment Heterogenity: Predicted treatment effects against their rank, standardized math test data. Average PTE's for each quartile are presented.

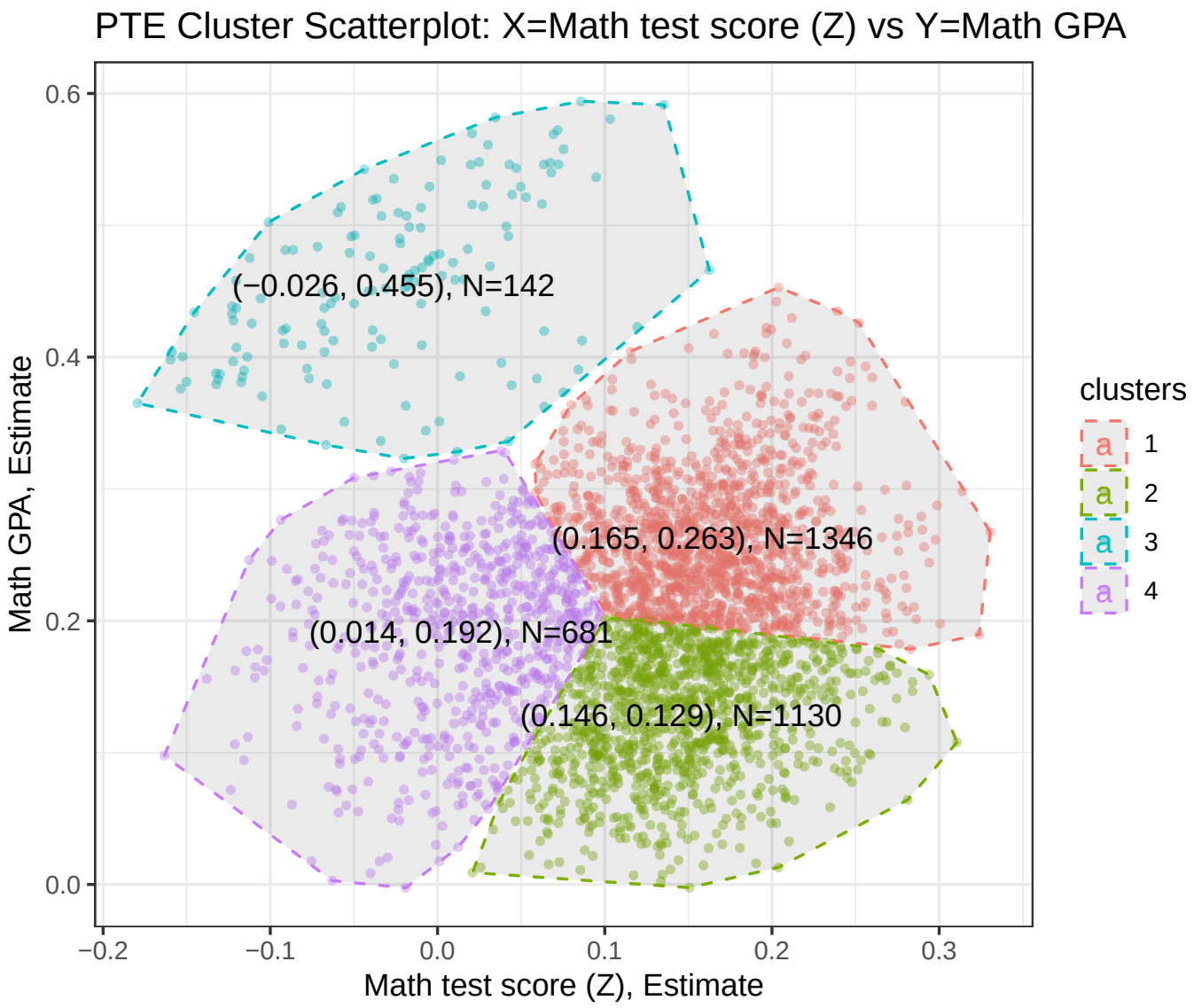

Figure 3: Random Forests for Treatment Heterogenity: Plotting PTEs of multiple outcomes - Math GPA vs Math test scores. Clusters estimated using k-means, k=4; k chosen through data-driven selection method.

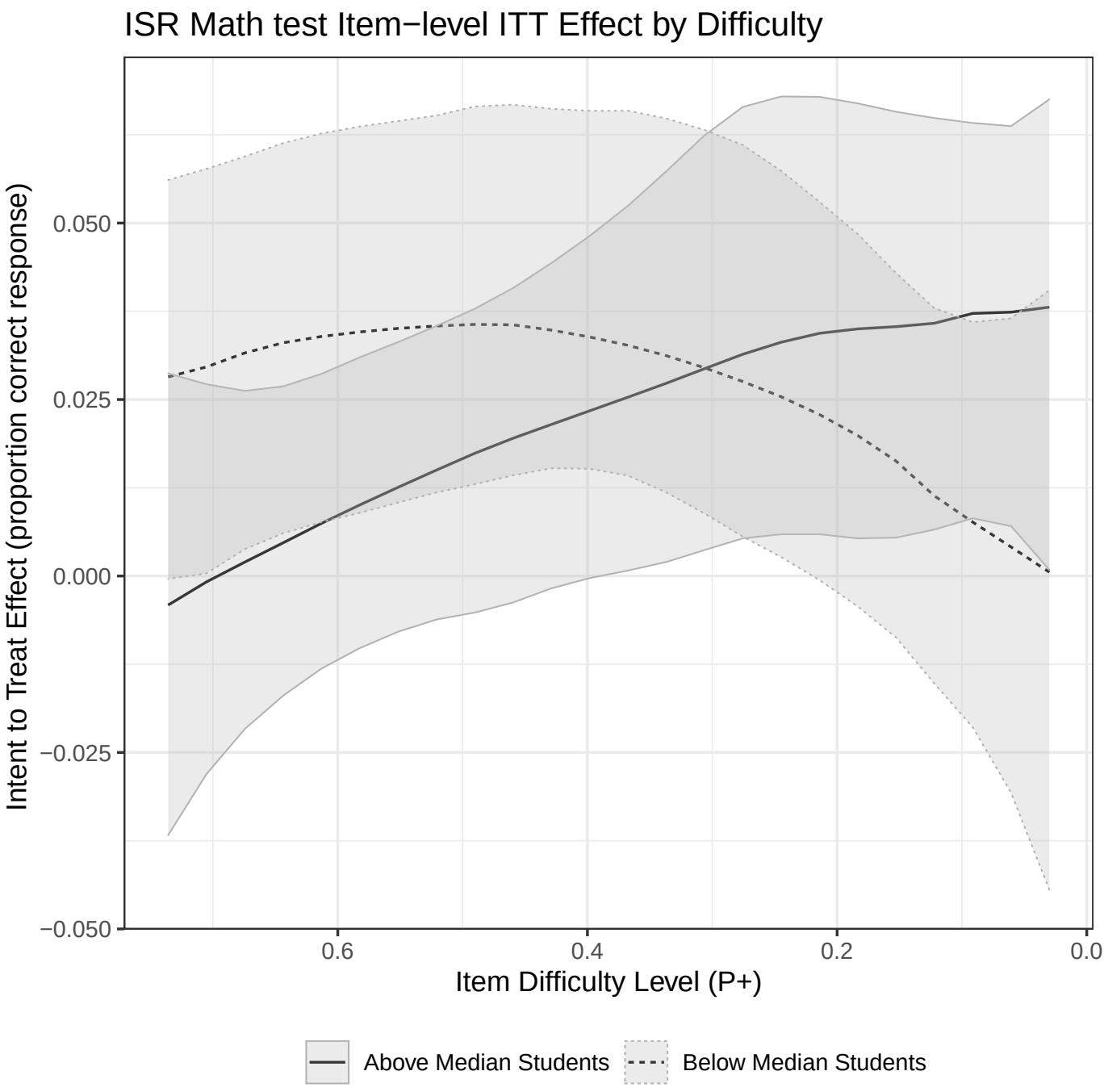

Figure 4: Treatment effects on individual ISR items against item-level difficulty for students above/below the median baseline standardized math test score. Above/Below median is calculated for all students taking the ISR math asessment. The lines shown are local polynomial fitted lines to see any underlying nonlinearities in the relationships between item-level impacts and difficulty. We also include [0.5, 0.95] confidence intervals on the fitted lines. This plot truncates the 3 easiest items (P-plus > 0.75) to better display the graph; the full version is included in the appendices.

1 Appendices

Table 1: (Appendix) Missing Data by Treatment Assignment: Study 1 and 2

	Control	Treatment
Study 1: Percent Missing		
N	1326	1307
Grade Data - Baseline year	6.49	7.88
Attendance Data - Baseline year	3.09	3.06
Standardized Math Test - Baseline year	14.48	16.68
F-test p-value: 0.997		
Grade Data - Program year 1	14.63	14.84
Attendance Data - Program year 1	4.37	6.2
Standardized Math Test - Program year 1	29.79	29.53
F-test p-value: 0.958		
Grade Data - Program year 2	27.98	27.54
Attendance Data - Program year 2	14.71	14.31
Standardized Math Test - Program year 2	37.41	38.03
F-test p-value: 0.981		
Study 2: Percent Missing		
N	1145	1565
Grade Data - Baseline Year	6.55	9.58
Attendance Data - Baseline Year	3.23	5.88
Standardized Math test - Baseline Year	9.96	12.14
F-test p-value: 0.946		
Grade Data - Study Year	23.67	20.89
Attendance Data - Study Year	8.91	8.56
Standardized Math test - Study Year	31.18	31.18
F-test p-value: 0.995		

Table 2: (Appendix) Estimated Effects on Academic and Behavioral Outcomes in Study 1, Year 1: Permutation Test

	N	Control Mean	Intent to Treat Estimate	Robust Standard Error	Permutation P-value
Mathematics					
SY 2014 Math Test Scaled Score (Z)	1852	-0.000	0.082**	(0.036)	0.252
SY 2014 Math GPA	2215	1.760	0.274***	(0.04)	0***
SY 2014 Math Courses Failed (percent)	2215	0.191	-0.043***	(0.013)	0.01***
ISR Math Test 2014 Z-Score	617	-0.000	0.119**	(0.058)	0.183
ISR Math Test 2014 Scaled Score	617	31.110	1.344**	(0.566)	0.128
Other Academics					
SY 2014 Read Test Scaled Score (Z)	1851	0.000	0.004	(0.04)	0.869
SY 2014 Non-Math GPA	2244	1.739	0.081**	(0.033)	0.781
SY 2014 Non-Math Core Courses Failed (percent)	2244	0.210	-0.027**	(0.011)	0.436
Behavior					
SY 2014 Disciplinary Incidents	2494	1.513	0.092	(0.104)	0.443
SY 2014 Days Absent	2494	24.242	0.7	(0.831)	0.226
SY 2014 Out-of-School Suspensions	2494	1.515	0.199	(0.153)	0.158
Crime					
SY 2014 Arrests for Violent Crimes	2633	0.086	-0.015	(0.015)	0.462
SY 2014 Arrests for Property Crimes	2633	0.061	-0.01	(0.01)	0.258
SY 2014 Arrests for Drug Crimes	2633	0.057	0.02	(0.014)	0.187
SY 2014 Arrests for Other Crimes	2633	0.178	-0.002	(0.022)	0.926
SY 2014 Ever Arrested	2633	0.176	-0.007	(0.013)	0.963
SY 2014 Total Arrests	2633	0.382	-0.008	(0.037)	0.994

Permutation tests were performed by randomly shuffling treatment assignment at the randomization block level and performing a (2-sided) t-test at each repetition. We then calculate the share of replications where this exceeds the t-test statistic using actual treatment assignment. This process is repeated for 100000 repetitions for each outcome.
Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
"Ever Arrested" indicates whether the student had any arrests during the period.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.

Table 3: (Appendix) Estimated Effects on Academic and Behavioral Outcomes in Study 2, Year 1: Permutation Test

	N	Control Mean	ITT Estimate	Clustered Standard Error	Permutation P-value
Mathematics					
SY 2015 Math Test Scaled Score (Z)	1865	0.008	0.126***	(0.036)	0.014**
SY 2015 Math GPA	2062	1.859	0.15***	(0.043)	0.012**
SY 2015 Math Courses Failed (percent)	2062	0.149	-0.029**	(0.013)	0.159
Other Academics					
SY 2015 Read Test Scaled Score (Z)	1865	0.007	-0.007	(0.04)	0.778
SY 2015 Non-Math GPA	2110	1.936	0.065*	(0.034)	0.283
SY 2015 Non-Math Core Courses Failed (percent)	2110	0.138	-0.011	(0.01)	0.628
Behavior					
SY 2015 Disciplinary Incidents	2474	1.554	-0.002	(0.138)	0.434
SY 2015 Days Absent	2474	22.777	0.756	(0.822)	0.701
SY 2015 Out-of-School Suspensions	2474	0.732	0.07	(0.091)	0.215
Crime					
SY 2015 Violent Crime Arrests	2710	0.104	-0.013	(0.016)	0.465
SY 2015 Property Crime Arrests	2710	0.072	-0.026	(0.016)	0.069*
SY 2015 Drug Crime Arrests	2710	0.051	0	(0.012)	0.635
SY 2015 Other Crime Arrests	2710	0.225	-0.052*	(0.028)	0.123
SY 2015 Ever Arrested	2710	0.164	-0.019	(0.013)	0.184
SY 2015 Total Arrests	2710	0.452	-0.091**	(0.045)	0.06*

Permutation tests were performed by randomly shuffling treatment assignment at the randomization block level and performing a (2-sided) t-test at each repetition. We then calculate the share of replications where this exceeds the t-test statistic using actual treatment assignment. This process is repeated for 100000 repetitions for each outcome.
Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
"Ever Arrested" indicates whether the student had any arrests during the period.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.

Table 4: (Appendix) Estimated 1 Year Effects on Academic and Behavioral Outcomes, Pooling Study 1 and 2: Permutation Test

	N	Control Mean	ITT Estimate	Clustered Standard Error	Permutation P-value
Mathematics					
Math Test Scaled Score (Z)	3717	0.004	0.113***	(0.026)	0.008***
Math GPA	4277	1.803	0.219***	(0.029)	0***
Math Courses Failed (percent)	4277	0.173	-0.037***	(0.009)	0.001***
Other Academics					
Read Test Scaled Score (Z)	3716	0.003	0	(0.028)	0.758
Non-Math GPA	4354	1.825	0.077***	(0.024)	0.093*
Non-Math Core Courses Failed (percent)	4354	0.178	-0.02***	(0.007)	0.066*
Behavior					
Disciplinary Incidents	4968	1.532	0.047	(0.086)	0.284
Days Absent	4968	23.581	0.767	(0.582)	0.313
Out-of-School Suspensions	4968	1.161	0.136	(0.089)	0.196
Crime					
Arrests for Violent Crimes	5343	0.094	-0.013	(0.011)	0.355
Arrests for Property Crimes	5343	0.066	-0.018*	(0.01)	0.035**
Arrests for Drug Crimes	5343	0.054	0.009	(0.009)	0.595
Arrests for Other Crimes	5343	0.200	-0.026	(0.018)	0.326
Ever Arrested	5343	0.171	-0.012	(0.009)	0.24
Total Arrests	5343	0.414	-0.048	(0.029)	0.169

Permutation tests were performed by randomly shuffling treatment assignment at the randomization block level and performing a (2-sided) t-test at each repetition. We then calculate the share of replications where this exceeds the t-test statistic using actual treatment assignment. This process is repeated for 100000 repetitions for each outcome.
Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
"Ever Arrested" indicates whether the student had any arrests during the period.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.

Table 5: (Appendix) Estimated Effects of Year 2 Treatment on Academic and Behavioral Outcomes in Study 1: Permutation Test

	N	Control Mean	ITT Estimate	Standard Error	Permutation P-value
Mathematics					
SY 2015 Math Test Scaled Score (Z)	1640	-0.000	0.163***	(0.038)	0.045**
SY 2015 Math GPA	1841	1.870	0.14***	(0.049)	0.441
SY 2015 Math Courses Failed (percent)	1841	0.168	-0.021	(0.015)	0.986
ISR Math Test 2015 Z-Score	878	0.000	0.067	(0.052)	0.756
ISR Math Test 2015 Scaled Score	878	33.178	0.984*	(0.563)	0.944
Other Academics					
SY 2015 Read Test Scaled Score (Z)	1640	0.000	-0.055	(0.044)	0.097*
SY 2015 Non-Math GPA	1895	1.860	0.035	(0.041)	0.942
SY 2015 Non-Math Core Courses Failed (percent)	1895	0.180	-0.007	(0.012)	0.884
Behavior					
SY 2015 Disciplinary Incidents	2251	1.126	-0.036	(0.105)	0.927
SY 2015 Days Absent	2251	23.141	-0.24	(0.899)	0.698
SY 2015 Out-of-School Suspensions	2251	0.716	0.049	(0.095)	0.617
Crime					
SY 2015 Arrests for Violent Crimes	2633	0.089	0	(0.015)	0.929
SY 2015 Arrests for Property Crimes	2633	0.057	0.002	(0.012)	0.959
SY 2015 Arrests for Drug Crimes	2633	0.084	0.025	(0.02)	0.202
SY 2015 Arrests for Other Crimes	2633	0.187	0.036	(0.028)	0.274
SY 2015 Ever Arrested	2633	0.159	0.019	(0.013)	0.161
SY 2015 Total Arrests	2633	0.418	0.063	(0.047)	0.275

Permutation tests were performed by randomly shuffling treatment assignment at the randomization block level and performing a (2-sided) t-test at each repetition. We then calculate the share of replications where this exceeds the t-test statistic using actual treatment assignment. This process is repeated for 100000 repetitions for each outcome.
Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
"Ever Arrested" indicates whether the student had any arrests during the period.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.

Table 6: (Appendix) 11th Grade Outcomes: Study 1

Outcome	N	Control Mean	Intent-to-Treat Effect	ITT Standard Error	Treatment-on-treated Effect	TOT Standard Error	Control Complier Mean	FDR Q-value
Held Back by 11th Grade	1722	0.119	0.004	(0.013)	0.008	(0.025)	0.100	0.756
Mathematics								
11th Grade Math Test (z-score)	1528	0.010	0.158***	(0.04)	0.3***	(0.074)	-0.214	0.001
11th Grade GPA: Math	1590	2.002	0.061	(0.052)	0.124	(0.097)	1.872	0.201
Non-math Academics								
11th Grade GPA: All Non-Math Courses	1721	1.963	-0.01	(0.042)	-0.011	(0.079)	1.853	0.890
11th Grade Reading Test (z-score)	1528	0.012	0.018	(0.044)	0.035	(0.081)	-0.189	0.890

Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above. Robust standard errors used.
To account for individuals having multiple observations, standard errors are clustered on individuals.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included. Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives.
Each panel in the above table is a 'family', so for example, 'Mathematics' represents one outcome family. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995).

Table 7: (Appendix) 11th Grade Outcomes: Study 2

Outcome	N	Control Mean	Intent-to-Treat Effect	ITT Standard Error	Treatment-on-treated Effect	TOT Standard Error	Control Complier Mean	FDR Q-value
Held Back by 11th Grade	1611	0.079	0.003	(0.013)	0.01	(0.04)	0.061	0.810
Mathematics								
11th Grade Math Test (z-score)	1445	0.000	0.017	(0.041)	0.046	(0.126)	0.046	0.717
11th Grade GPA: Math	1496	1.953	0.079	(0.052)	0.244	(0.156)	1.809	0.239
Non-math Academics								
11th Grade GPA: All Non-Math Courses	1618	1.920	0.047	(0.043)	0.146	(0.132)	1.721	0.534
11th Grade Reading Test (z-score)	1444	0.001	-0.005	(0.043)	-0.021	(0.133)	-0.008	0.875

Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above.
To account for individuals having multiple observations, standard errors are clustered on individuals.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included. Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives.
Each panel in the above table is a 'family', so for example, 'Mathematics' represents one outcome family. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995).

Table 8: (Appendix) Graduation outcomes: Study 1

Outcome	N	Control Mean	Intent-to-Treat Effect	ITT Standard Error	Treatment-on-treated Effect	TOT Standard Error	Control Complier Mean	FDR Q-value
Graduated On-Time	1853	0.738	0.008	(0.017)	0.017	(0.036)	0.759	0.74
Ever Graduated	1859	0.826	0.005	(0.015)	0.011	(0.033)	0.862	0.74
Leave code: Dropout	2728	0.062	0.005	(0.01)	0.013	(0.024)	0.052	0.74
Leave code: Corrections	2728	0.032	0.005	(0.007)	0.012	(0.017)	0.010	0.74
Leave code: Transfer	2728	0.195	0.006	(0.015)	0.014	(0.037)	0.120	0.74
Leave code: Unknown	2728	0.125	-0.011	(0.012)	-0.027	(0.029)	0.118	0.74
Leave code: No exit	2728	0.015	-0.009**	(0.004)	-0.021**	(0.01)	0.027	0.27
Leave code: Deceased	2728	0.009	-0.003	(0.003)	-0.006	(0.008)	0.012	0.74

Robust standard errors are used and presented in the models above.
To account for individuals having multiple observations, standard errors are clustered on individuals.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives.
We treat all outcomes above as one 'family'. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995).

Table 9: (Appendix) Graduation outcomes: Study 2

Outcome	N	Control Mean	Intent-to-Treat Effect	ITT Standard Error	Treatment-on-treated Effect	TOT Standard Error	Control Complier Mean	FDR Q-value
Graduated On-Time	1897	0.761	0.008	(0.019)	0.025	(0.06)	0.788	0.919
Ever Graduated	1911	0.821	0.003	(0.017)	0.006	(0.055)	0.857	0.919
Leave code: Dropout	2935	0.044	-0.002	(0.008)	-0.005	(0.028)	0.034	0.919
Leave code: Corrections	2935	0.023	0.001	(0.007)	0.005	(0.022)	0.011	0.919
Leave code: Transfer	2935	0.219	-0.003	(0.017)	-0.012	(0.055)	0.172	0.919
Leave code: Unknown	2935	0.145	-0.013	(0.014)	-0.046	(0.045)	0.191	0.826
Leave code: No exit	2935	0.037	0.011	(0.007)	0.035	(0.025)	0.013	0.740
Leave code: Deceased	2935	0.010	-0.005	(0.004)	-0.018	(0.013)	0.020	0.740

Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above.
To account for individuals having multiple observations, standard errors are clustered on individuals.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives.
We treat all outcomes above as one 'family'. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995).

Table 10: (Appendix) All Graduation/Enrollment outcomes: Pooling Studies 1 and 2

Outcome	N	Control Mean	Intent-to-Treat Effect	ITT Standard Error	Treatment-on-treated Effect	TOT Standard Error	Control Complier Mean
Graduated On-Time	3750	0.749	0.007	(0.013)	0.019	(0.034)	0.772
Graduated On-Time Post-Match	3755	0.757	0.006	(0.013)	0.014	(0.034)	0.783
Ever Graduated	3770	0.823	0.004	(0.012)	0.008	(0.032)	0.859
Graduated On-Time (zeros)	5653	0.492	0.008	(0.013)	0.023	(0.036)	0.558
Graduated On-Time Post-Match (zeros)	5662	0.499	0.007	(0.013)	0.021	(0.036)	0.564
Ever Graduated (zeros)	5663	0.543	0.009	(0.013)	0.027	(0.037)	0.610
Ever Graduated (Transfers/Unknowns=Dropouts)	5663	0.543	0.009	(0.013)	0.027	(0.037)	0.610
Ever Graduated (Transfers/Unknowns=Graduates)	5663	0.884	-0.002	(0.01)	-0.006	(0.028)	0.908
Leave code: Dropout	5663	0.054	0.002	(0.007)	0.006	(0.019)	0.042
Leave code: Corrections	5663	0.028	0.003	(0.006)	0.009	(0.017)	0.011
Leave code: Transfer	5663	0.207	0.002	(0.012)	0.005	(0.033)	0.143
Leave code: Unknown	5663	0.134	-0.013	(0.01)	-0.038	(0.027)	0.154
Leave code: No exit	5663	0.026	0.001	(0.004)	0.003	(0.012)	0.023
Leave code: Deceased	5663	0.009	-0.004	(0.003)	-0.012	(0.009)	0.016

Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above.
To account for individuals having multiple observations, standard errors are clustered on individuals.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives.
We treat all outcomes above as one 'family'. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995).

Table 11: (Appendix) Sensitivity of ITT Estimates to Choice of Baseline Covariates (Pooled Data from Study 1 and 2)

	All Covariates		Sociodemographic		Academic		Criminal Activity		No Covariates	
	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
Mathematics										
Math GPA	0.219***	(0.029)	0.188***	(0.034)	0.226***	(0.029)	0.202***	(0.034)	0.198***	(0.035)
Math Courses Failed (percent)	-0.037***	(0.009)	-0.032***	(0.01)	-0.039***	(0.009)	-0.035***	(0.01)	-0.035***	(0.01)
Math Test Scaled Score (Z)	0.113***	(0.026)	0.089***	(0.031)	0.114***	(0.026)	0.099***	(0.034)	0.095***	(0.034)
Other Academics										
Read Test Scaled Score (Z)	0	(0.028)	-0.015	(0.033)	-0.002	(0.029)	-0.011	(0.034)	-0.014	(0.034)
Non-Math GPA	0.077***	(0.024)	0.051*	(0.029)	0.08***	(0.024)	0.061**	(0.029)	0.059**	(0.03)
Non-Math Core Courses Failed (percent)	-0.02***	(0.007)	-0.014*	(0.008)	-0.021***	(0.007)	-0.018**	(0.008)	-0.017**	(0.009)
Behavior										
Disciplinary Incidents	0.047	(0.086)	0.065	(0.092)	0.031	(0.091)	0.047	(0.091)	0.043	(0.093)
Days Absent	0.767	(0.582)	0.948	(0.669)	0.618	(0.586)	0.79	(0.66)	0.722	(0.686)
Out-of-School Suspensions	0.136	(0.089)	0.154	(0.097)	0.127	(0.096)	0.149	(0.094)	0.139	(0.098)
Crime										
Arrests for Violent Crimes	-0.013	(0.011)	-0.011	(0.011)	-0.014	(0.011)	-0.012	(0.011)	-0.012	(0.011)
Arrests for Property Crimes	-0.018*	(0.01)	-0.016	(0.01)	-0.019*	(0.01)	-0.018*	(0.01)	-0.018*	(0.01)
Arrests for Drug Crimes	0.009	(0.009)	0.009	(0.01)	0.006	(0.01)	0.009	(0.009)	0.007	(0.01)
Arrests for Other Crimes	-0.026	(0.018)	-0.024	(0.02)	-0.032*	(0.019)	-0.026	(0.018)	-0.03	(0.02)
Ever Arrested	-0.012	(0.009)	-0.008	(0.01)	-0.014	(0.01)	-0.01	(0.009)	-0.012	(0.01)
Total Arrests	-0.048	(0.029)	-0.042	(0.033)	-0.059*	(0.032)	-0.046	(0.03)	-0.053	(0.034)

Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above.
Sections indicate different combinations of baseline covariates used in ITT models.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included.
Sociodemographic: race, gender, age, learning disability, subsidized lunch, baseline grade; Academic: GPA, absences, grades earned; Criminal Activity: arrests.
Math Scaled Score refers to the respective CPS standardized test administered in a given year to a given grade level. See text for details.
Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
"Ever Arrested" indicates whether the student had any arrests during the period.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.

Table 12: (Appendix) Estimated Effects on Academic and Behavioral Outcomes in Study 1, Year 1 (No Shows Omitted)

	N	Control Mean	Intent to Treat Estimate	Robust Standard Error	Effect of Treatment on Treated (TOT)	Robust Standard Error	Control Complier Mean	False Discovery Rate Q-Value
Mathematics								
SY 2014 Math Test Scaled Score (Z)	1597	0.007	0.086**	(0.039)	0.147**	(0.064)	-0.077	0.029
SY 2014 Math GPA	1991	1.765	0.279***	(0.041)	0.522***	(0.075)	1.665	0.001
SY 2014 Math Courses Failed (percent)	1991	0.186	-0.038***	(0.013)	-0.07***	(0.024)	0.162	0.007
Other Academics								
SY 2014 Read Test Scaled Score (Z)	1596	0.004	-0.013	(0.043)	-0.019	(0.071)	-0.085	0.788
SY 2014 Non-Math GPA	2011	1.739	0.074**	(0.034)	0.161***	(0.059)	1.763	0.021
SY 2014 Non-Math Core Courses Failed (percent)	2011	0.206	-0.023**	(0.011)	-0.043**	(0.021)	0.206	0.063
Behavior								
SY 2014 Disciplinary Incidents	2115	1.559	0.093	(0.105)	0.19	(0.205)	1.486	0.598
SY 2014 Days Absent	2115	24.913	0.433	(0.898)	0.911	(1.737)	23.407	0.598
SY 2014 Out-of-School Suspensions	2115	1.645	0.105	(0.166)	0.211	(0.323)	1.749	0.598
Crime								
SY 2014 Arrests for Violent Crimes	2129	0.092	-0.017	(0.017)	-0.034	(0.033)	0.104	0.585
SY 2014 Arrests for Property Crimes	2129	0.064	-0.015	(0.012)	-0.03	(0.023)	0.061	0.585
SY 2014 Arrests for Drug Crimes	2129	0.058	0.017	(0.016)	0.034	(0.031)	0.011	0.585
SY 2014 Arrests for Other Crimes	2129	0.177	0.011	(0.025)	0.016	(0.048)	0.101	0.862
SY 2014 Ever Arrested	2129	0.183	-0.01	(0.014)	-0.021	(0.028)	0.152	0.686
SY 2014 Total Arrests	2129	0.392	-0.005	(0.042)	-0.014	(0.081)	0.276	0.862

Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included.
"No Show" indicates that the student did not enroll in Fall of study start in the school they were expected to attend.
Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
"Ever Arrested" indicates whether the student had any arrests during the period.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives. Each panel in the above table is a 'family', so for example, 'Mathematics' represents one outcome family. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995)

Table 13: (Appendix) Estimated Effects on Academic and Behavioral Outcomes in Study 2, Year 1 (No Shows Omitted)

	N	Control Mean	ITT Estimate	Clustered Standard Error	Effect of Treatment on Treated (TOT)	Clustered Standard Error	Control Complier Mean	False Discovery Rate Q-Value
Mathematics								
SY 2015 Math Test Scaled Score (Z)	1416	-0.018	0.145***	(0.043)	0.359***	(0.107)	-0.100	0.002
SY 2015 Math GPA	1703	1.869	0.168***	(0.046)	0.431***	(0.12)	1.840	0.002
SY 2015 Math Courses Failed (percent)	1703	0.147	-0.034**	(0.014)	-0.087**	(0.035)	0.181	0.013
Other Academics								
SY 2015 Read Test Scaled Score (Z)	1415	-0.060	0.009	(0.046)	0.033	(0.113)	-0.105	0.768
SY 2015 Non-Math GPA	1730	1.950	0.053	(0.037)	0.136	(0.096)	2.031	0.355
SY 2015 Non-Math Core Courses Failed (percent)	1730	0.139	-0.012	(0.01)	-0.031	(0.026)	0.153	0.355
Behavior								
SY 2015 Disciplinary Incidents	1844	1.635	-0.073	(0.15)	-0.243	(0.406)	1.860	0.877
SY 2015 Days Absent	1844	24.433	0.241	(0.953)	0.421	(2.561)	24.013	0.877
SY 2015 Out-of-School Suspensions	1844	0.861	0.024	(0.113)	0.048	(0.307)	0.708	0.877
Crime								
SY 2015 Violent Crime Arrests	1848	0.115	-0.015	(0.02)	-0.041	(0.055)	0.143	0.538
SY 2015 Property Crime Arrests	1848	0.089	-0.033	(0.02)	-0.089	(0.054)	0.122	0.163
SY 2015 Drug Crime Arrests	1848	0.054	-0.003	(0.016)	-0.008	(0.042)	0.037	0.849
SY 2015 Other Crime Arrests	1848	0.250	-0.051	(0.033)	-0.144	(0.09)	0.286	0.163
SY 2015 Ever Arrested	1848	0.191	-0.029*	(0.016)	-0.079*	(0.043)	0.196	0.163
SY 2015 Total Arrests	1848	0.507	-0.101*	(0.054)	-0.283*	(0.146)	0.587	0.163

Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above.
To account for individuals having multiple observations, standard errors are clustered on individuals.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included.
"No Show" indicates that the student did not enroll in Fall of study start in the school they were expected to attend.
Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
"Ever Arrested" indicates whether the student had any arrests during the period.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives. Each panel in the above table is a 'family', so for example, 'Mathematics' represents one outcome family. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995)

Table 14: (Appendix) Study 1: Variations on Missing Outcome Data Imputation

			Original Results		Quantile Regression (QR)		Multiple Imputation (MI)	
	N	Control Mean	Intent to Treat Estimate	Robust Standard Error	QR Estimate	QR Standard Error	MI Estimate	MI Standard Error
SY 2014 Math Test Scaled Score (Z)	1852	0.000	0.082**	(0.036)	0.067***	(0.023)	0.079*	(0.045)
SY 2014 Math GPA	2215	1.760	0.274***	(0.04)	0.28***	(0.061)	0.268***	(0.04)
SY 2014 Math Courses Failed (percent)	2215	0.191	-0.043***	(0.013)			-0.04***	(0.013)
SY 2014 Read Test Scaled Score (Z)	1851	0.000	0.004	(0.04)	-0.004	(0.025)	-0.006	(0.042)
SY 2014 Non-Math GPA	2244	1.739	0.081**	(0.033)	0.073	(0.045)	0.084**	(0.033)
SY 2014 Non-Math Core Courses Failed (percent)	2244	0.210	-0.027**	(0.011)			-0.026**	(0.011)
SY 2014 Days Absent	2494	24.242	0.7	(0.831)	-0.213	(0.525)	0.702	(0.82)

We present our standard results alongside different approaches to imputing missing data.

We run median quantile regression after imputing 0's for the outcome variables, denoted as 'QR'. We calculate bootstrap standard errors.

We also perform multiple imputation via chained equations (denoted 'MI'). We impute M=50 datasets and pool the estimated effects and robust standard errors.

Table 15: (Appendix) Study 2: Variations on Missing Outcome Data Imputation

			Original Results		Quantile Regression (QR)		Multiple Imputation (MI)	
	N	Control Mean	ITT Estimate	Clustered Standard Error	QR Estimate	QR Standard Error	MI Estimate	MI Standard Error
SY 2015 Math Test Scaled Score (Z)	1865	0.008	0.126***	(0.036)	0.074***	(0.02)	0.085*	(0.045)
SY 2015 Math GPA	2062	1.859	0.15***	(0.043)	0.206***	(0.053)	0.107**	(0.043)
SY 2015 Math Courses Failed (percent)	2062	0.149	-0.029**	(0.013)			-0.021	(0.013)
SY 2015 Read Test Scaled Score (Z)	1865	0.007	-0.007	(0.04)	-0.016	(0.02)	-0.024	(0.046)
SY 2015 Non-Math GPA	2110	1.936	0.065*	(0.034)	0.107**	(0.047)	0.049	(0.036)
SY 2015 Non-Math Core Courses Failed (percent)	2110	0.138	-0.011	(0.01)			-0.009	(0.01)
SY 2015 Days Absent	2474	22.777	0.756	(0.822)	0.171	(0.47)	0.85	(0.813)

We present our standard results alongside different approaches to imputing missing data.

We run median quantile regression after imputing 0's for the outcome variables, denoted as 'QR'. We calculate bootstrap standard errors.

We also perform multiple imputation via chained equations (denoted 'MI'). We impute M=50 datasets and pool the estimated effects and robust standard errors.

N=65 were randomized into Study 2 twice.

Both assignments are retained in the models above and standard errors are clustered at the student level.

Table 16: (Appendix) Estimated 1 Year Treatment Effects: Pooling Study 1 and 2 - 9th Grade Student Subsample Only

	N	Control Mean	ITT Estimate	Clustered Standard Error	Effect of Treatment on Treated (TOT)	Clustered Standard Error	Control Complier Mean	False Discovery Rate Q-Value
Mathematics								
Math Test Scaled Score (Z)	2735	0.006	0.102***	(0.03)	0.263***	(0.077)	-0.063	0.002
Math GPA	3020	1.832	0.158***	(0.035)	0.391***	(0.087)	1.801	0.001
Math Courses Failed (percent)	3020	0.165	-0.025**	(0.011)	-0.062**	(0.026)	0.161	0.019
Other Academics								
Read Test Scaled Score (Z)	2734	0.001	0.003	(0.033)	0.017	(0.084)	-0.090	0.836
Non-Math GPA	3083	1.875	0.057**	(0.029)	0.144**	(0.072)	1.933	0.138
Non-Math Core Courses Failed (percent)	3083	0.165	-0.013	(0.009)	-0.034	(0.021)	0.171	0.164
Behavior								
Disciplinary Incidents	3579	1.440	0.083	(0.099)	0.23	(0.274)	1.419	0.402
Days Absent	3579	22.817	0.974	(0.677)	2.57	(1.88)	21.170	0.402
Out-of-School Suspensions	3579	1.028	0.099	(0.097)	0.269	(0.268)	0.978	0.402
Crime								
Arrests for Violent Crimes	3905	0.092	-0.009	(0.013)	-0.026	(0.038)	0.109	0.607
Arrests for Property Crimes	3905	0.072	-0.021*	(0.012)	-0.063*	(0.035)	0.096	0.123
Arrests for Drug Crimes	3905	0.053	0.005	(0.011)	0.015	(0.032)	0.020	0.629
Arrests for Other Crimes	3905	0.212	-0.06***	(0.02)	-0.168***	(0.061)	0.317	0.036
Ever Arrested	3905	0.165	-0.019*	(0.01)	-0.053*	(0.03)	0.182	0.123
Total Arrests	3905	0.429	-0.085**	(0.034)	-0.241**	(0.102)	0.541	0.054

Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above.
To account for individuals having multiple observations, standard errors are clustered on individuals.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included.
Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
"Ever Arrested" indicates whether the student had any arrests during the period.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives. Each panel in the above table is a 'family', so for example, 'Mathematics' represents one outcome family. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995)

Table 17: (Appendix) Estimated 1 Year Treatment Effects: Pooling Study 1 and 2 - 10th Grade Student Subsample Only

	N	Control Mean	ITT Estimate	Clustered Standard Error	Effect of Treatment on Treated (TOT)	Clustered Standard Error	Control Complier Mean	False Discovery Rate Q-Value
Mathematics								
Math Test Scaled Score (Z)	976	-0.002	0.127**	(0.05)	0.251**	(0.098)	-0.254	0.011
Math GPA	1235	1.750	0.328***	(0.052)	0.709***	(0.112)	1.487	0.001
Math Courses Failed (percent)	1235	0.189	-0.049***	(0.017)	-0.107***	(0.038)	0.202	0.007
Other Academics								
Read Test Scaled Score (Z)	976	0.009	-0.012	(0.057)	-0.025	(0.111)	-0.097	0.823
Non-Math GPA	1248	1.714	0.087**	(0.042)	0.189**	(0.092)	1.659	0.114
Non-Math Core Courses Failed (percent)	1248	0.209	-0.025*	(0.014)	-0.055*	(0.031)	0.240	0.114
Behavior								
Disciplinary Incidents	1340	1.778	-0.174	(0.18)	-0.403	(0.418)	2.311	0.699
Days Absent	1340	25.195	0.451	(1.135)	1.018	(2.614)	26.241	0.699
Out-of-School Suspensions	1340	1.503	0.081	(0.207)	0.184	(0.476)	1.670	0.699
Crime								
Arrests for Violent Crimes	1371	0.100	-0.027	(0.021)	-0.064	(0.049)	0.143	0.559
Arrests for Property Crimes	1371	0.053	-0.008	(0.017)	-0.018	(0.041)	0.079	0.788
Arrests for Drug Crimes	1371	0.055	0.017	(0.018)	0.038	(0.041)	0.018	0.703
Arrests for Other Crimes	1371	0.161	0.056	(0.036)	0.129	(0.084)	0.011	0.559
Ever Arrested	1371	0.187	0.001	(0.02)	0.002	(0.046)	0.146	0.968
Total Arrests	1371	0.370	0.038	(0.055)	0.085	(0.128)	0.251	0.757

Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above.
To account for individuals having multiple observations, standard errors are clustered on individuals.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included.
Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
"Ever Arrested" indicates whether the student had any arrests during the period.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives. Each panel in the above table is a 'family', so for example, 'Mathematics' represents one outcome family. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995)

Table 18: (Appendix) Year 1 Effect Estimates - Female Students Only

	N	Control Mean	ITT Estimate	Clustered Standard Error	Effect of Treatment on Treated (TOT)	Clustered Standard Error	Control Complier Mean	False Discovery Rate Q-Value
Mathematics								
Math Test Scaled Score (Z)	595	-0.054	0.12**	(0.059)	0.389**	(0.191)	-0.136	0.064
Math GPA	658	2.007	0.177**	(0.073)	0.556**	(0.234)	1.774	0.054
Math Courses Failed (percent)	658	0.117	-0.021	(0.02)	-0.069	(0.062)	0.180	0.267
Other Academics								
Read Test Scaled Score (Z)	596	0.083	0.048	(0.067)	0.153	(0.218)	-0.107	0.675
Non-Math GPA	678	2.163	0.064	(0.059)	0.2	(0.19)	2.012	0.675
Non-Math Core Courses Failed (percent)	678	0.109	-0.008	(0.016)	-0.021	(0.05)	0.138	0.675
Behavior								
Disciplinary Incidents	767	1.700	-0.303	(0.227)	-1.061	(0.799)	2.875	0.554
Days Absent	767	25.549	-1.091	(1.449)	-3.544	(4.968)	32.176	0.564
Out-of-School Suspensions	767	0.636	0.09	(0.147)	0.289	(0.499)	0.306	0.564
Crime								
Arrests for Violent Crimes	817	0.065	-0.004	(0.021)	-0.013	(0.075)	0.066	0.862
Arrests for Property Crimes	817	0.031	-0.014	(0.017)	-0.05	(0.059)	0.082	0.483
Arrests for Drug Crimes	817	0.006	0.013	(0.01)	0.053	(0.036)	-0.029	0.204
Arrests for Other Crimes	817	0.108	-0.091***	(0.034)	-0.318***	(0.123)	0.340	0.059
Ever Arrested	817	0.105	-0.032	(0.02)	-0.113	(0.071)	0.188	0.204
Total Arrests	817	0.210	-0.096**	(0.048)	-0.328*	(0.169)	0.459	0.161

Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above.
To account for individuals having multiple observations, standard errors are clustered on individuals.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included.
Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
"Ever Arrested" indicates whether the student had any arrests during the period.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives. Each panel in the above table is a 'family', so for example, 'Mathematics' represents one outcome family. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995)

Table 19: (Appendix) Year 1 Effects Estimates - Male Students Only (pooling studies 1 and 2)

	N	Control Mean	ITT Estimate	Clustered Standard Error	Effect of Treatment on Treated (TOT)	Clustered Standard Error	Control Complier Mean	False Discovery Rate Q-Value
Mathematics								
Math Test Scaled Score (Z)	3122	0.013	0.11***	(0.029)	0.25***	(0.064)	-0.126	0.001
Math GPA	3619	1.769	0.232***	(0.032)	0.526***	(0.072)	1.647	0.001
Math Courses Failed (percent)	3619	0.182	-0.04***	(0.01)	-0.09***	(0.023)	0.186	0.001
Other Academics								
Read Test Scaled Score (Z)	3120	-0.010	-0.009	(0.032)	-0.014	(0.071)	-0.090	0.840
Non-Math GPA	3676	1.769	0.077***	(0.026)	0.178***	(0.06)	1.804	0.010
Non-Math Core Courses Failed (percent)	3676	0.190	-0.021***	(0.008)	-0.05***	(0.019)	0.204	0.013
Behavior								
Disciplinary Incidents	4201	1.504	0.081	(0.091)	0.208	(0.232)	1.492	0.369
Days Absent	4201	23.257	1.006	(0.637)	2.478	(1.624)	21.680	0.290
Out-of-School Suspensions	4201	1.248	0.135	(0.103)	0.342	(0.263)	1.191	0.290
Crime								
Arrests for Violent Crimes	4526	0.099	-0.015	(0.012)	-0.04	(0.033)	0.126	0.522
Arrests for Property Crimes	4526	0.072	-0.018	(0.011)	-0.05	(0.031)	0.092	0.522
Arrests for Drug Crimes	4526	0.062	0.008	(0.01)	0.023	(0.028)	0.021	0.522
Arrests for Other Crimes	4526	0.215	-0.015	(0.02)	-0.036	(0.055)	0.200	0.522
Ever Arrested	4526	0.182	-0.008	(0.01)	-0.02	(0.027)	0.162	0.522
Total Arrests	4526	0.449	-0.04	(0.033)	-0.102	(0.091)	0.439	0.522

Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above.
To account for individuals having multiple observations, standard errors are clustered on individuals.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included.
Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
"Ever Arrested" indicates whether the student had any arrests during the period.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives. Each panel in the above table is a 'family', so for example, 'Mathematics' represents one outcome family. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995)

Table 20: (Appendix) ITT Estimates Interacted with Whether School Above Median in Arrests per-capita During School Year (Pooled)

Year 1 Outcome	N	Ctrl Mean	Intent-to-Treat Est.	ITT Standard Error	Dummy x Treatment Status	Interaction SE
Mathematics						
Math GPA	4331	1.786	0.306***	(0.045)	-0.144**	(0.060)
Math Courses Failed (percent)	4331	0.179	-0.055***	(0.013)	0.028	(0.018)
Standardized Math Score	3741	0.001	0.164***	(0.040)	-0.085	(0.053)
Non-math academics						
Non-Math GPA	4409	1.809	0.087**	(0.038)	-0.018	(0.050)
Non-math Core Courses Failed (percent)	4409	0.184	-0.028**	(0.011)	0.016	(0.015)
Standardized Reading Score	3740	0.001	0	(0.044)	0.004	(0.058)
Behavior						
Disciplinary incidents	5035	1.549	-0.213**	(0.092)	0.414**	(0.162)
Days Absent	5035	23.932	-0.647	(0.709)	2.656**	(1.122)
Out-of-School Incs	5035	1.180	-0.117	(0.085)	0.429**	(0.167)
Crime						
Violent Crime Arrests	5410	0.095	-0.018	(0.014)	0.01	(0.020)
Property Crime Arrests	5410	0.073	-0.004	(0.009)	-0.028	(0.018)
Drug Arrests	5410	0.054	-0.007	(0.008)	0.03*	(0.016)
Other arrests	5410	0.207	-0.041**	(0.018)	0.023	(0.033)
Any Arrests (Dummy)	5410	0.176	-0.029**	(0.013)	0.03*	(0.018)
All Arrests	5410	0.429	-0.071**	(0.031)	0.035	(0.055)

Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above. To account for individuals having multiple observations, standard errors are clustered on individuals. Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included. Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science). * = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.

Given that randomization block fixed-effects fully explain school fixed-effects, we cannot estimate the main effect for the 'above median' dummy variable, so we do not report this in the table. However, we can recover and report the interaction effect.

Table 21: (Appendix) Intent-to-treat (ITT) Estimates: Black and Latinx Subsample, pooling both studies

Outcome	N	Ctrl Mean	ITT	ITT Standard Error	ITT Q-val	ITT x Latinx Interaction	ITT x Latinx Standard Error	ITT x Latinx Q-val	ITT Joint Test P-val	ITT Joint Test Q-val
Academics: Math										
Math GPA	4133	1.784	0.196***	(0.041)	0.001	-0.004	(0.061)	0.946	0.006***	0.010
Math Courses Failed (percent)	4133	0.178	-0.032**	(0.014)	0.018	-0.002	(0.019)	0.946	0.28	0.280
Math Test Scaled Score (Z)	3574	-0.008	0.084**	(0.035)	0.018	0.029	(0.052)	0.946	0.001***	0.003
Academics: Non-math										
Non-Math Core GPA	4209	1.797	0.092***	(0.034)	0.019	-0.053	(0.049)	0.663	0.757	0.940
Non-Math Core Courses Failed (percent)	4209	0.184	-0.022**	(0.011)	0.060	0.012	(0.015)	0.663	0.444	0.940
Read Test Scaled Score (Z)	3573	-0.009	-0.002	(0.039)	0.953	-0.018	(0.058)	0.754	0.94	0.940
Behavior										
Disciplinary Incidents	4814	1.589	0.001	(0.138)	0.993	0.111	(0.169)	0.876	0***	0.001
Days Absent	4814	24.139	1.121	(0.896)	0.321	-0.189	(1.208)	0.876	0.101	0.102
Out-of-School Suspensions	4814	1.204	0.182	(0.147)	0.321	-0.06	(0.175)	0.876	0***	0.001
Crime										
Arrests for Violent Crimes	5162	0.098	-0.008	(0.018)	0.671	-0.011	(0.02)	0.620	0***	0.002
Arrests for Property Crimes	5162	0.070	-0.023	(0.017)	0.330	0.013	(0.02)	0.620	0.099*	0.149
Arrests for Drug Crimes	5162	0.056	0.007	(0.015)	0.671	0.009	(0.017)	0.620	0.935	0.936
Arrests for Other Crimes	5162	0.213	-0.053*	(0.028)	0.320	0.056	(0.037)	0.620	0.879	0.936
Ever Arrested	5162	0.179	-0.016	(0.014)	0.391	0.013	(0.018)	0.620	0.001***	0.002
Total Arrests	5162	0.437	-0.076	(0.047)	0.320	0.065	(0.058)	0.620	0.038**	0.077

We test for differences between the Black and Latinx students in our study sample by running our normal ITT/TOT analysis, and including a term interacting treatment with an indicator variable for 'Latinx'. We present those estimates above, along with corresponding FDR Q-values. We also test whether the coefficients on the interaction term and the Latinx indicator variable are jointly 0; we present this p-value and corresponding FDR Q-value in the table above. Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above. To account for individuals having multiple observations, standard errors are clustered on individuals.
The compliance rate for Black students is 0.35 and the compliance rate for Latinx students is 0.43.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included. Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives.
Each panel in the above table is a 'family', so for example, 'Mathematics' represents one outcome family. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995).

Table 22: (Appendix) Treatment-on-the-treated (TOT) Estimates: Black and Latinx Subsample, pooling both studies

Outcome	N	Ctrl Mean	TOT	TOT Standard Error	TOT Q-val	TOT x Latinx Interaction	TOT x Latinx Standard Error	TOT x Latinx Q-val	TOT Joint Test P-val	TOT Joint Test Q-val
Academics: Math										
Math GPA	4133	1.784	0.538***	(0.114)	0.001	-0.143	(0.147)	0.900	0.531	0.798
Math Courses Failed (percent)	4133	0.178	-0.089**	(0.038)	0.020	0.017	(0.047)	0.900	0.817	0.818
Math Test Scaled Score (Z)	3574	-0.008	0.241**	(0.101)	0.020	-0.016	(0.126)	0.900	0.162	0.487
Academics: Non-math										
Non-Math Core GPA	4209	1.797	0.254***	(0.094)	0.021	-0.169	(0.121)	0.463	0.403	0.605
Non-Math Core Courses Failed (percent)	4209	0.184	-0.06**	(0.03)	0.063	0.038	(0.038)	0.463	0.263	0.605
Read Test Scaled Score (Z)	3573	-0.009	0.004	(0.111)	0.975	-0.041	(0.139)	0.771	0.812	0.813
Behavior										
Disciplinary Incidents	4814	1.589	0.003	(0.425)	0.994	0.244	(0.48)	0.687	0.066*	0.100
Days Absent	4814	24.139	3.374	(2.771)	0.336	-1.335	(3.31)	0.687	0.314	0.314
Out-of-School Suspensions	4814	1.204	0.557	(0.453)	0.336	-0.283	(0.502)	0.687	0.054*	0.100
Crime										
Arrests for Violent Crimes	5162	0.098	-0.023	(0.058)	0.690	-0.02	(0.061)	0.783	0.07*	0.422
Arrests for Property Crimes	5162	0.070	-0.075	(0.055)	0.336	0.049	(0.059)	0.614	0.931	0.932
Arrests for Drug Crimes	5162	0.056	0.025	(0.049)	0.690	0.015	(0.053)	0.783	0.884	0.932
Arrests for Other Crimes	5162	0.213	-0.17*	(0.092)	0.336	0.186*	(0.107)	0.489	0.175	0.524
Ever Arrested	5162	0.179	-0.051	(0.046)	0.406	0.046	(0.053)	0.614	0.521	0.932
Total Arrests	5162	0.437	-0.245	(0.155)	0.336	0.23	(0.174)	0.560	0.831	0.932

We test for differences between the Black and Latinx students in our study sample by running our normal ITT/TOT analysis, and including a term interacting treatment with an indicator variable for 'Latinx'. We present those estimates above, along with corresponding FDR Q-values. We also test whether the coefficients on the interaction term and the Latinx indicator variable are jointly 0; we present this p-value and corresponding FDR Q-value in the table above. Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above. To account for individuals having multiple observations, standard errors are clustered on individuals.
The compliance rate for Black students is 0.35 and the compliance rate for Latinx students is 0.43.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free and reduced lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included. Non-math GPA refers to all non math courses in core subject areas (English, Science, Social Science).
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives.
Each panel in the above table is a 'family', so for example, 'Mathematics' represents one outcome family. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995).

Table 23: (Appendix) Variations on Calculating Grade Variables: Pooling results from both studies

Outcome	N	Control Mean	Intent-to-Treat Effect	ITT Standard Error	Treatment-on-treated Effect	TOT SE	Control Complier Mean	FDR Q-value
Math Outcomes								
Number of Math Courses	4331	1.900	0.007	(0.009)	0.017	(0.021)	1.942	0.423
Math GPA	4331	1.786	0.219***	(0.03)	0.521***	(0.071)	1.661	0.001
Math Course Failures	4331	0.308	-0.07***	(0.017)	-0.166***	(0.04)	0.348	0.001
Math Course Failures (percent)	4331	0.179	-0.038***	(0.01)	-0.091***	(0.023)	0.191	0.001
Math GPA (no selections)	4210	1.807	0.225***	(0.03)	0.525***	(0.071)	1.677	0.001
Math Course Failures (no selections)	4210	0.297	-0.07***	(0.017)	-0.163***	(0.039)	0.337	0.001
Math Course Failures (percent) (no selections)	4210	0.172	-0.038***	(0.01)	-0.089***	(0.022)	0.185	0.001
Non-Math Core: All classes								
Number of Non-math Core Courses	4409	6.712	-0.061	(0.044)	-0.144	(0.105)	6.905	0.210
Non-math Core GPA	4409	1.809	0.076***	(0.024)	0.182***	(0.059)	1.689	0.019
Non-math Core Course Failures	4409	1.112	-0.113**	(0.05)	-0.271**	(0.121)	1.296	0.051
Non-math Core Course Failures (percent)	4409	0.184	-0.019**	(0.008)	-0.045**	(0.018)	0.200	0.032
Non-Math Core: High grade by subject								
Number of Non-math Core Courses	4409	5.518	0.024	(0.029)	0.061	(0.069)	5.679	0.400
Non-math Core GPA	4409	1.864	0.071***	(0.025)	0.17***	(0.059)	1.751	0.025
Non-math Core Course Failures	4409	0.838	-0.076**	(0.038)	-0.184**	(0.092)	0.951	0.062
Non-math Core Course Failures (percent)	4409	0.170	-0.015**	(0.007)	-0.037**	(0.018)	0.180	0.051
Non-Math Core: Low grade by subject								
Number of Non-math Core Courses	4409	5.518	0.024	(0.029)	0.061	(0.069)	5.679	0.400
Non-math Core GPA	4409	1.746	0.079***	(0.024)	0.191***	(0.059)	1.626	0.019
Non-math Core Course Failures	4409	0.970	-0.103**	(0.041)	-0.248**	(0.098)	1.140	0.032
Non-math Core Course Failures (percent)	4409	0.196	-0.021***	(0.008)	-0.052***	(0.019)	0.216	0.025
Non-Math Core: Top 3 classes each semester								
Number of Non-math Core Courses	4409	5.656	0.023	(0.027)	0.057	(0.065)	5.793	0.400
Non-math Core GPA	4409	1.916	0.067***	(0.025)	0.161***	(0.06)	1.805	0.025
Non-math Core Course Failures	4409	0.800	-0.082**	(0.039)	-0.197**	(0.093)	0.919	0.051
Non-math Core Course Failures (percent)	4409	0.160	-0.016**	(0.007)	-0.038**	(0.018)	0.171	0.051
Non-Math Core: Top 6 classes in that year								
Number of Non-math Core Courses	4409	5.713	0.016	(0.024)	0.041	(0.058)	5.835	0.479
Non-math Core GPA	4409	1.927	0.069***	(0.025)	0.165***	(0.061)	1.816	0.025
Non-math Core Course Failures	4409	0.810	-0.088**	(0.039)	-0.211**	(0.094)	0.916	0.051
Non-math Core Course Failures (percent)	4409	0.158	-0.016**	(0.007)	-0.039**	(0.018)	0.167	0.051

We report variations on the calculation of grade variables (GPA and course failures) for both Math and non-Math courses.
For math, our preferred method is to select the appropriate grade-level math grade for each student as their main 'math course';
e.g. for a 9th grader, if they're in only one math course that semester, we take that math grade; if they are in multiple courses that semester and take an algebra course (the grade-level course for 9th graders), we 'pick' that algebra grade; if they are in multiple courses and take geometry (the 10th grade level course), we take that grade. If they don't take either, then we randomly select a math class.
For non-math core courses, we present 5 methods:
- 'All classes' includes all non-math core (science, english, and social science) classes
- 'High grade by subject' selects the highest grade for each non-math core topic in each semester, and uses those courses to calculate the outcomes.
- 'Low grade by subject' is the same procedure as above, but uses the lowest grade for each non-math core subject per semester.
- 'Top 3 classes each semester' uses a student's three highest non-math grades in a semester regardless of subject.
- 'Top 6 classes in that year' uses a student's six highest non-math grades in that school year, regardless of subject.
Some students (N=65) were randomized into Study 2 twice. Both assignments are retained in the models above.
To account for individuals having multiple observations, standard errors are clustered on individuals.
Baseline covariates: Randomization block, treatment assignment, gender, age, learning disability, free lunch recipient, race, baseline grade level, GPA, days absent from school, disciplinary incidents, including suspensions, and arrests. Where missing data was found, zeroes were imputed and variables identifying records as missing data were created and included. Non-math GPA refers to all non math courses, not just core subjects.
* = p-value < 0.1, ** = p-value < 0.05, *** = p-value < 0.01.
False discovery rate (FDR) Q-values are the share of estimates within a family of related outcomes that are expected to be false positives.
All Math outcomes are one family and all non-Math Core outcomes are another family. We report the smallest FDR q-value at which we could reject the null for each outcome using the method from Benjamini and Hochberg (1995).

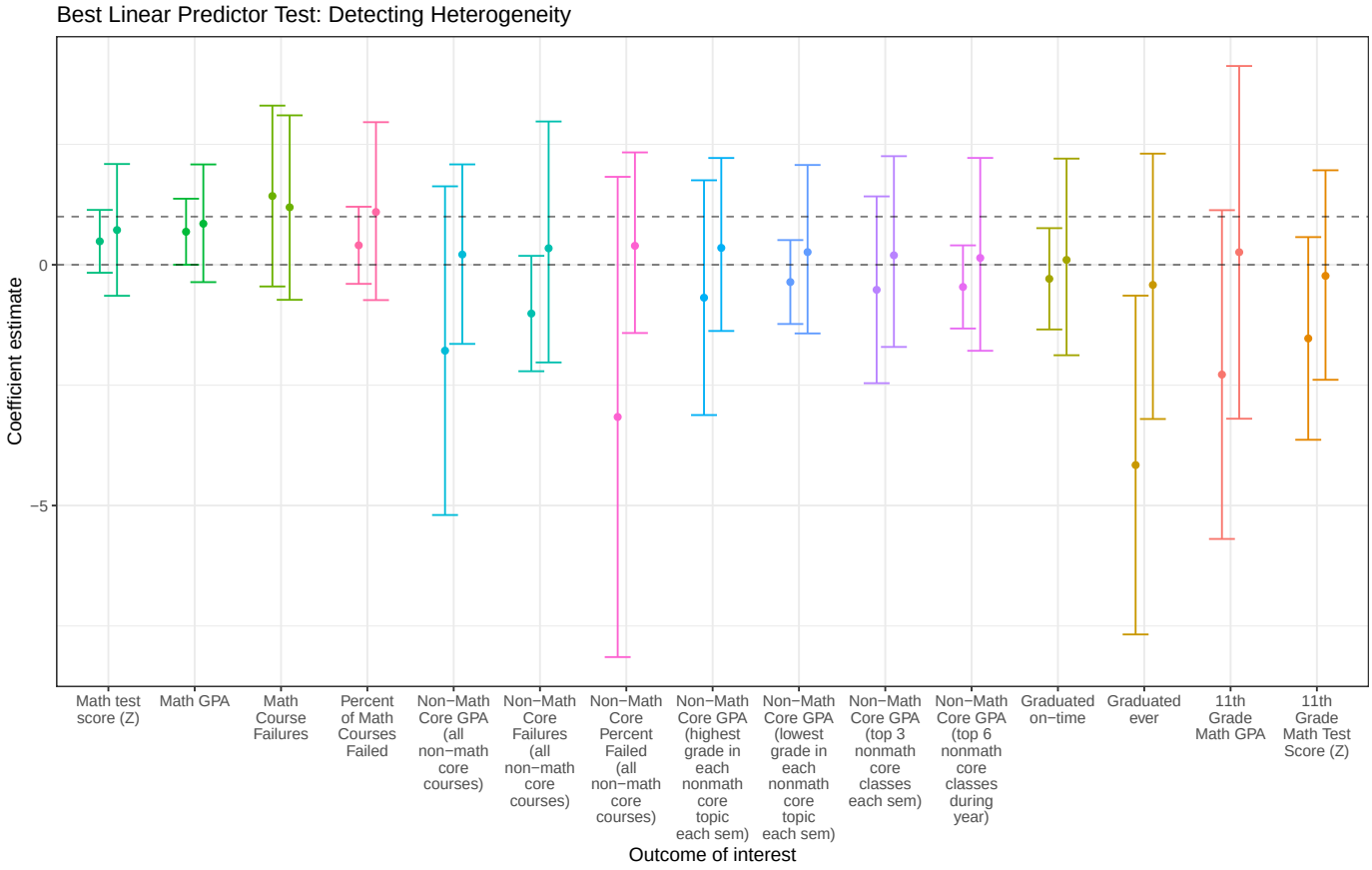

Figure 1: Random Forests for Treatment Heterogeneity - Best Linear Predictor test

For each outcome, the estimate and confidence intervals for the differential forest prediction coefficient of the BLP test are shown for the full-sample procedure (left) and split-sample procedure (right). Horizontal lines are plotted at 1 and 0.

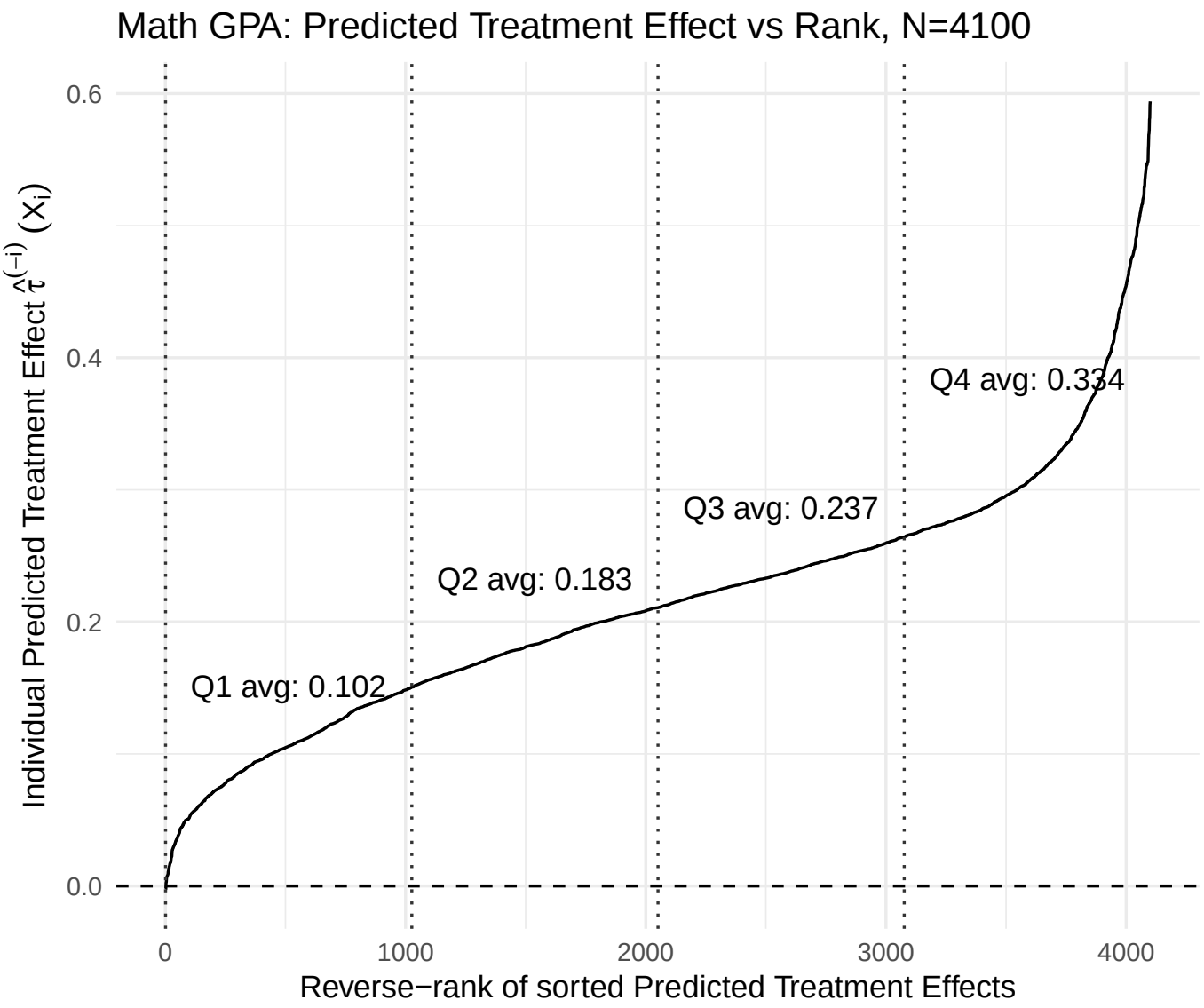

Figure 2: Random Forests for Treatment Heterogenity: Predicted treatment effects against their rank, math GPA. Average PTE's for each quartile are presented.

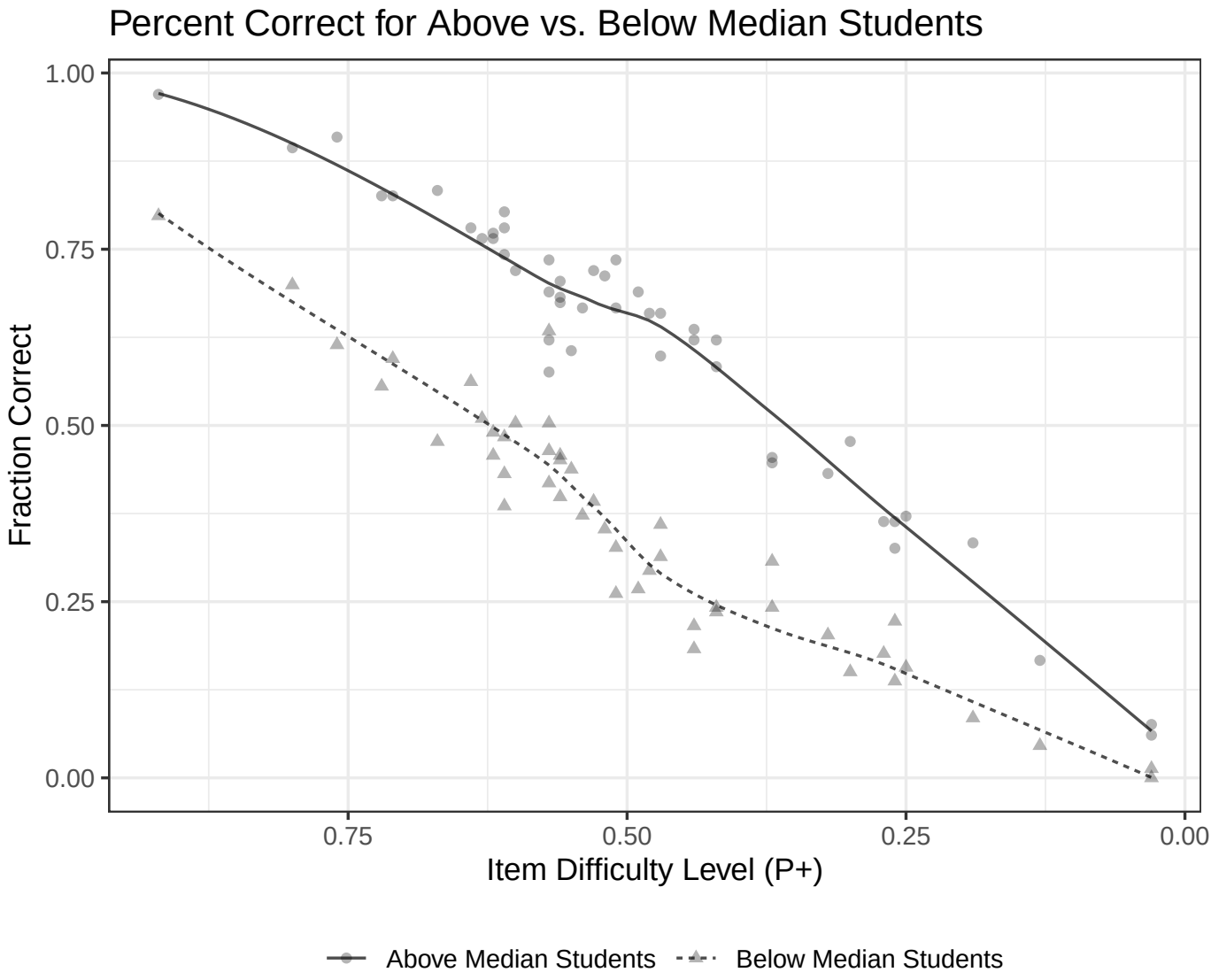

Figure 3: Percent of students correctly answering math items given on the ISR math assessment. Each point corresponds to a single item.

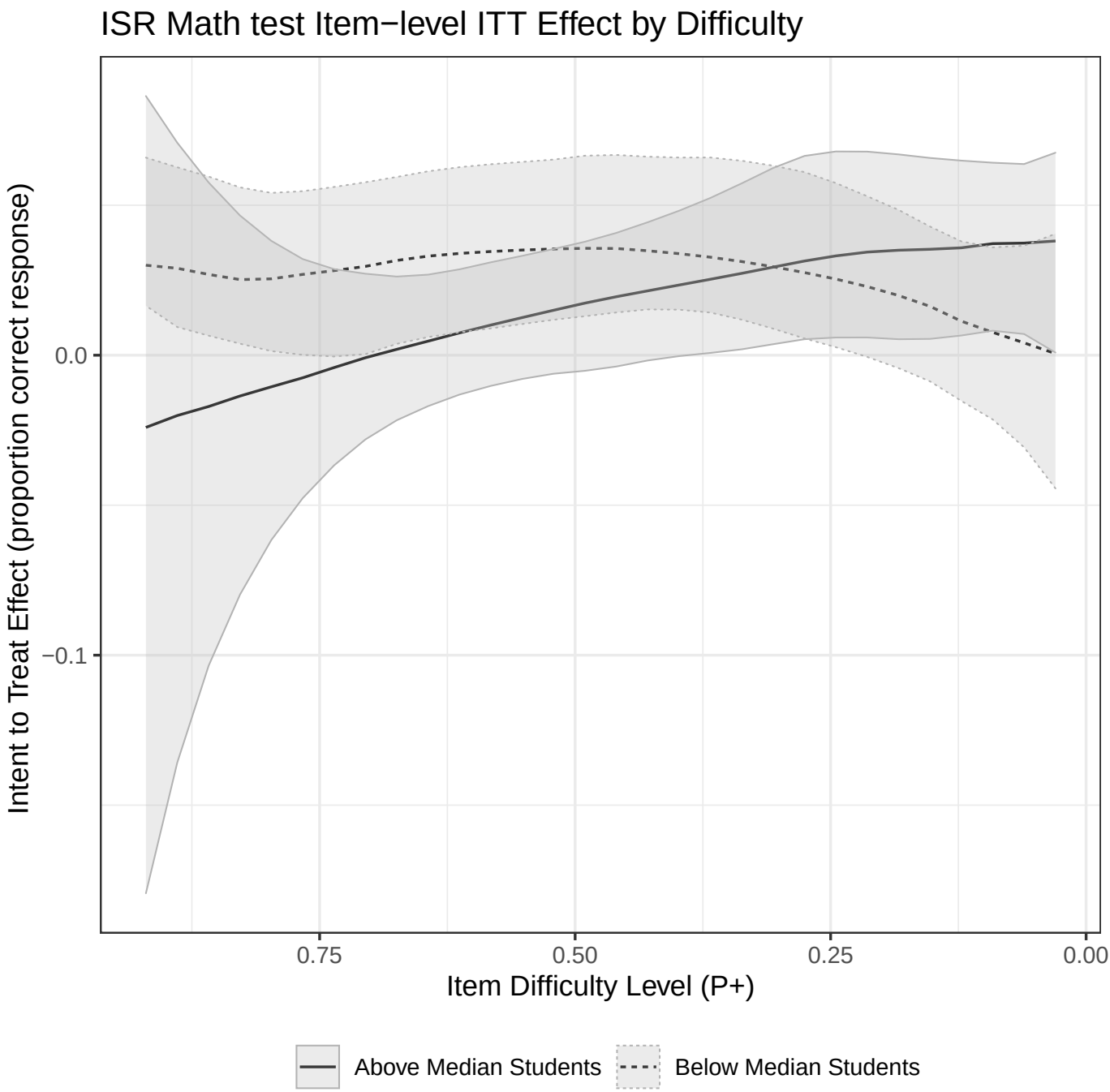

Figure 4: Treatment effects on individual ISR items against item-level difficulty for students above/below the median baseline standardized math test score. Above/Below median is calculated for all students taking the ISR math asessment. The lines shown are local polynomial fitted lines to see any underlying nonlinearities in the relationships between item-level impacts and difficulty. We also include [0.5, 0.95] confidence intervals on the fitted lines.