School Accountability, Test Scores, and Long-Run Outcomes

Joshua Hollinger

University of Rochester

November 10, 2021

School Accountability

- How to improve public schools?
  - Focus on inputs?
  - Measure and incentivize output?
- **No Child Left Behind Act of 2001**:
  Federal policy: school accountability for test scores
  - Literature suggests positive effects on test scores
  - Could be "teaching to the test", or insufficient for meaningful long-run returns
  - **Long-run effects?**
- This paper → estimate effects of NCLB on elementary students' test scores and end-of-high-school outcomes

This Paper

- Administrative data from North Carolina: grades 3-5
- NCLB: escalating sanctions for schools failing to make Adequate Yearly Progress (AYP)
  - Based on fraction of students proficient on tests
  - Target set for whole school and 9 subgroups
- 2 different RDDs to capture variation in accountability pressure
  - Subgroup RDD: school not accountable for subgroups with < 40 students
  - School RDD: sanctions start after second AYP failure $\rightarrow$ pressure after failing first time

Preview of Results

1. Elementary test score increases
2. Higher SAT scores in high school
3. Mixed effects on HS GPA, 4-year college intention
4. No effect on HS graduation, intention to attend any college
5. Persistent effects on subsequent test scores

Suggests NCLB led to improvement in math and reading skills in the long-run, but may not have improved educational attainment

Literature on NCLB

- Modest positive effect of NCLB on test scores (Chakrabarti 2014; Ahn & Vigdor 2014; Dee & Jacob 2011)
- **Effect of NCLB on long-run outcomes?** → this paper
- Evidence of "gaming" under school accountability:
  - Higher-calorie lunches (Figlio, Winicki 2005)
  - Less focus on low-stakes subjects (Jacob 2005)
  - High- vs. low-stakes exams (Koretz 2002)
- **Did "teaching to the test" harm students in the long-run?**
- Little evidence on the long-run effects of incentives for educators to increase test scores
  - Performance pay in Israel (Lavy 2020)
  - Accountability in Texas before NCLB (Deming et al. 2016)

Roadmap

1. Conceptual Framework
2. No Child Left Behind
3. Data
4. Empirical Strategy 1: Subgroup RDD
5. Empirical Strategy 2: School RDD
6. Results
7. Relationship between Test Scores and Long-Run Outcomes
8. Mechanisms and Heterogeneous Effects
9. Conclusion

Conceptual Framework - Multi-tasking

- Principal-agent model with multi-tasking (Holmstrom and Milgrom 1991)
- Teacher can focus on test-scores or broader student learning (2 tasks differ in test and long-run returns)
- Government only observes test scores and attaches incentive
- Suppose test scores increase as a result
- Long-run effects?
  - Could be **negative** if caused by substitution toward teaching to the test and long-run returns much lower when doing so
  - Could be **positive** if caused by increase in effort level and teaching to the test has similar long-run returns

No Child Left Behind (NCLB)

- NCLB: US federal accountability implemented 2002/03
- Proficiency test score was set for all students
- Adequate Yearly Progress (AYP) target requires a certain percentage of students in each school and subgroups within schools to be proficient
  $\rightarrow$ White, Black, Hispanic, Asian, Native American, Multi-racial, Economically Disadvantaged, Students with Disabilities, Limited English Proficiency
- Each objective must be passed, or the school fails AYP

No Child Left Behind (NCLB)

- Fail AYP twice consecutively $\rightarrow$ "school identified for improvement"
- Escalating sanctions for these schools with each following year of sanctions
- Must pass AYP twice consecutively to be removed from "improvement" list
- A school does not have to meet the proficiency target for subgroups with fewer than 40 students in NC
- Confidence interval and safe harbor exceptions

No Child Left Behind (NCLB)

- Policy features suggest several dimensions of variation in incentive strength
  1. Schools with fewer than 40 students in a subgroup have much less incentive to focus on those students
  2. Schools failing once are under threat of dealing with sanctions if they fail again

NCLB Sanctions (Ahn and Vigdor 2014)

- **Year 2: offer transfers to better schools in the district**
- Year 3: offer supplemental education services to free/reduced-price lunch students
- Year 4: "corrective actions" - include staff/leadership changes, curriculum changes, instructional time changes, or appointment of outside advisors
- Year 5: restructuring plan
- Year 6: restructure - conversion to a charter school, replacement of the principal and most staff, state takeover, contracting with another entity to manage the school

Data

- North Carolina Education Research Data Center
- All 3rd-5th grade students in NC 2003-2008 (NCLB started allowing retests in 2009), and their end of high school outcomes in 2010-2017
- Students' end-of-grade test scores and demographic characteristics
- Identifiers allow tracking students over time, matching to classrooms, teachers, and schools
- Merged with collected school-level AYP reports for each year

Empirical Strategies: Overview

- Subgroup RDD: subgroups with < 40 within a school not counted for accountability
  - Treatment group: students in a subgroup just at or above 40 in a school
  - Control group: students in the same subgroup in a different school with just below 40 students
- School RDD: schools that fail AYP a 1st time have strong incentive to improve to avoid sanctions accompanying 2nd failure
  - Treatment group: students in a school that barely failed AYP in prior year
  - Control group: students in a school that did not fail AYP in prior year

Summary Statistics

	Full Sample	**Subgroup RD Sample**	**School RD Sample**
Math score	0.021	-0.183	-0.040
Reading score	0.014	-0.213	-0.046
White	0.563	0.297	0.526
Black	0.274	0.378	0.317
Hispanic	0.084	0.237	0.088
Asian	0.021	0.042	0.020
Economically Disadvantaged	0.553	0.828	0.586
English Language Learner	0.053	0.143	0.057
Female	0.494	0.495	0.493
SAT-taker	0.345	0.281	0.348
SAT score	993.9	938.1	976.8
High School GPA	2.788	2.619	2.746
Dropout	0.049	0.060	0.058
Graduate	0.833	0.809	0.824
Intend 4-year college	0.453	0.380	0.448
Intend any college	0.852	0.816	0.852
School size	309.7	244.8	272.8
Student-year observations	1,304,301	56,246	184,004
School-year observations	7,714	1,349	1,135

Comparing Empirical Strategies

- No easy way to capture effect of NCLB accountability pressure
- Subgroup RD:
  - Variation in pressure specific to a subgroup of students
  - These students are lower-achieving on average
- School RD:
  - Variation in pressure at the school level
  - These are lower-achieving schools, but students are not as low achieving due to within-school variance
- LATE:
  - Accountability could have different effects on the samples close to RD cutoffs
  - Total accountability pressure could be greater than the variation captured by either RD

Empirical Strategy 1: Subgroup RDD

- School is only accountable for a subgroup of students if there are at least 40 students of that subgroup in the school that year

$$y_{isgt} = \tau T_{sgt} + \theta X_{sgt} + \phi(X_{sgt} \times T_{sgt}) + \lambda_g + \delta_t + \beta Z_{isgt} + \epsilon_{isgt}$$

i=student, s=school, g=subgroup, t=year

- X_{sgt} is running variable, T_{sgt} is indicator for above cutoff
- Bandwidth of 5
- Only students in the subgroup near 40 included
- Controls will include lag test scores and demographics
- Similar to Gilraine (2018)

Identification

- "As good as random" which side of 40 students a subgroup lands on
- Variation: comparing subgroups in schools with just below 40 students to same subgroup in another school with just over 40 students
- Only students in the school most of the year count for NCLB
- Also need to control for differences as we move away from the cutoff
- Test for bunching and discontinuity of covariates across cutoff

Empirical Strategy 2: School RDD

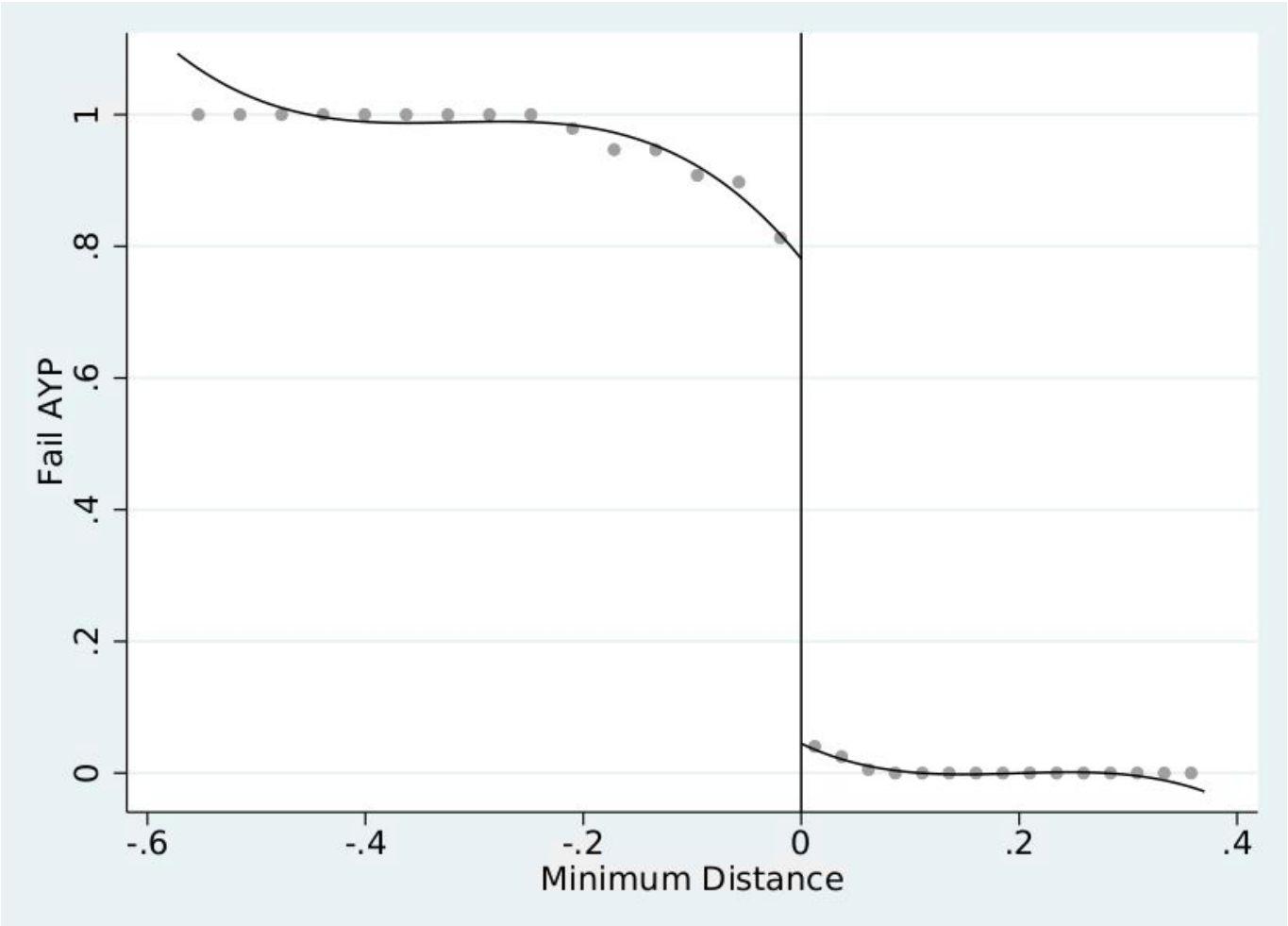
- Compare schools that barely passed AYP last year to those who barely failed last year
- MSE-optimal bandwidth (around 0.08)
- Restrict to those who didn't fail prior to last year: isolate first NCLB sanction exposure
- More commonly used in the literature (Chakrabarti 2014, Ahn and Vigdor 2014)

$$y_{ist} = \tau F_{s,t-1} + \theta M_{s,t-1} + \phi(M_{s,t-1} \times F_{s,t-1}) + \beta Z_{ist} + \epsilon_{ist}$$

Running variable: minimum distance

- Each school has many AYP objectives and must hit them all to avoid failing AYP
- Minimum distance defined as where the school is relative to the target on their worst-performing objective
- Target is minimum of required fraction proficient for confidence interval and safe harbor

First stage: Minimum distance and AYP Fail



Effect of Accountability Pressure on Test Scores

Subgroup RDD

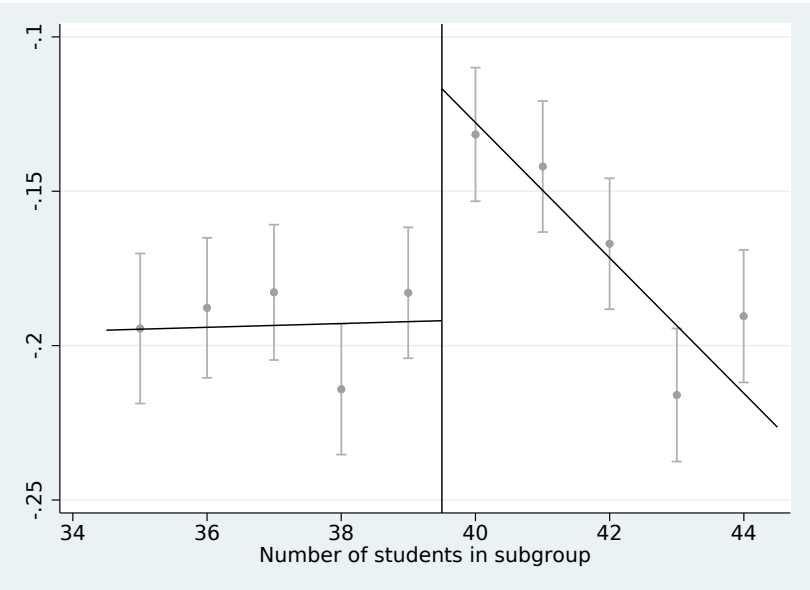
	Math		Reading	
Treated (τ_{sgt})	0.0814***	0.0615***	0.0625***	0.0389***
	(0.0158)	(0.0110)	(0.0180)	(0.00799)
Observations	50,702	50,702	50,814	50,814

School RDD

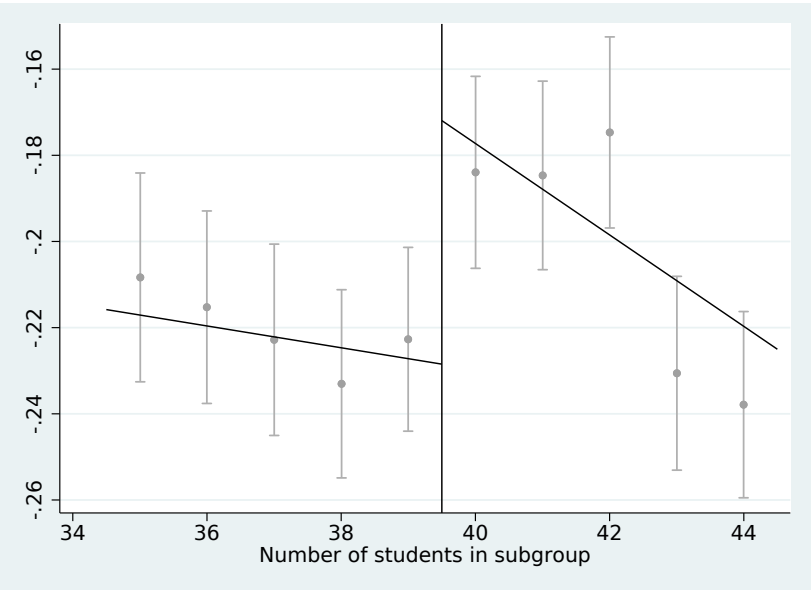
	Math		Reading	
Fail ($F_{s,t-1}$)	0.0907**	0.0493**	0.107***	0.0424***
	(0.0416)	(0.0247)	(0.0342)	(0.0162)
Observations	165,859	199,679	175,057	238,880
Student controls	NO	YES	NO	YES

Subgroup RDD - Test Score Plots

Math

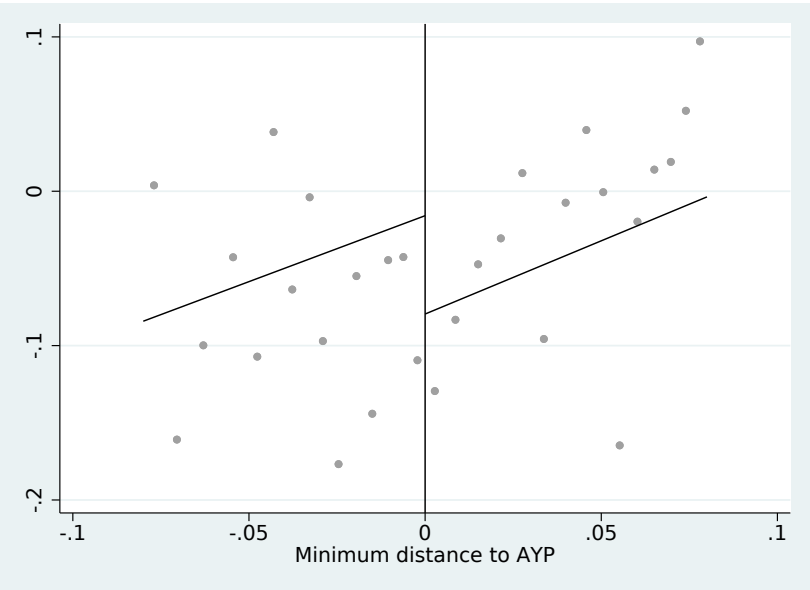


Reading

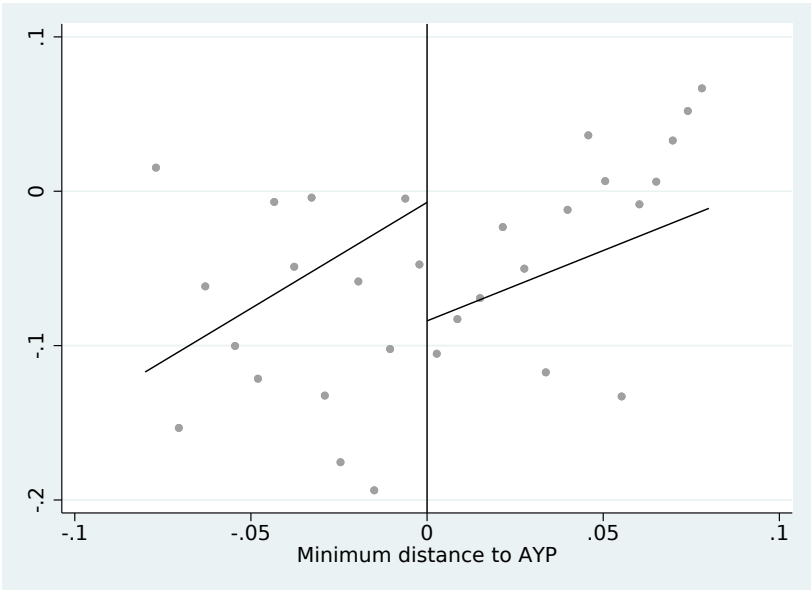


School RDD - Test Score Plots

Math



Reading



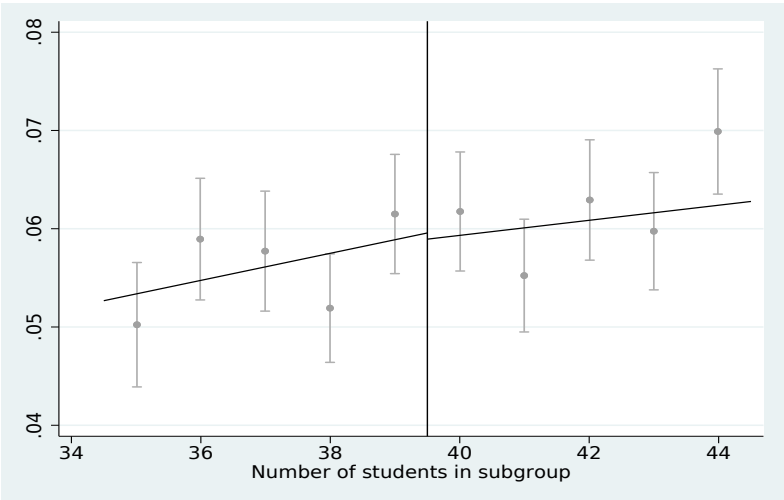
Subgroup RDD - Effects on Long-Run Outcomes

	Dropout		Graduate		HS GPA		SAT score	
Treated	-0.00132	-0.00103	-0.0053	-0.00522	0.0308***	0.0224**	19.67**	17.99***
	(0.00521)	(0.00509)	(0.00559)	(0.00538)	(0.0113)	(0.0104)	(8.508)	(6.212)
Mean Dep. Var.	0.049		0.833		2.788		993.9	
SD Dep. Var.	0.217		0.373		0.729		192.4	
Observations	41,055	41,055	41,055	41,055	27,308	27,308	14,295	14,295

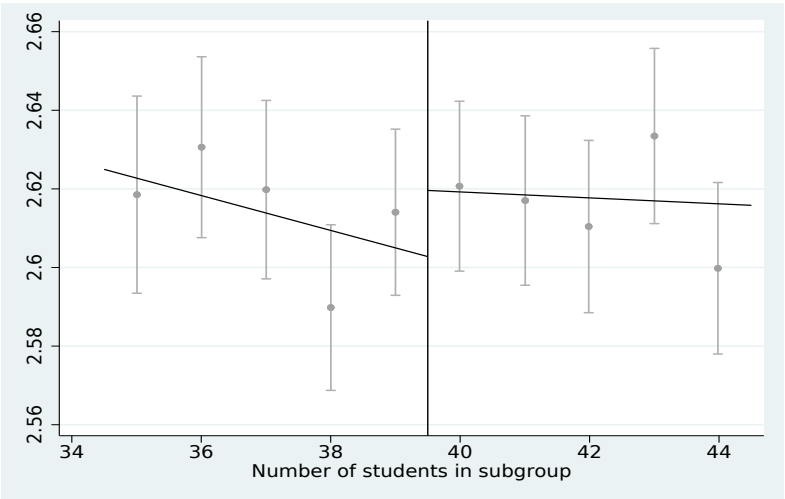
	SAT-taking		Intend college		Intend 4-year		Intend 2-year	
Treated	0.00591	0.00341	-0.00594	-0.00711	0.00328	-0.00233	-0.00921	-0.00478
	(0.00851)	(0.00844)	(0.00552)	(0.00551)	(0.00948)	(0.00706)	(0.00996)	(0.00849)
Mean Dep. Var.	0.345		0.852		0.453		0.399	
SD Dep. Var.	0.475		0.355		0.498		0.490	
Observations	50,747	50,747	33,230	33,230	33,230	33,230	33,230	33,230
Student controls	NO	YES	NO	YES	NO	YES	NO	YES

Subgroup RDD - Long-Run Plots

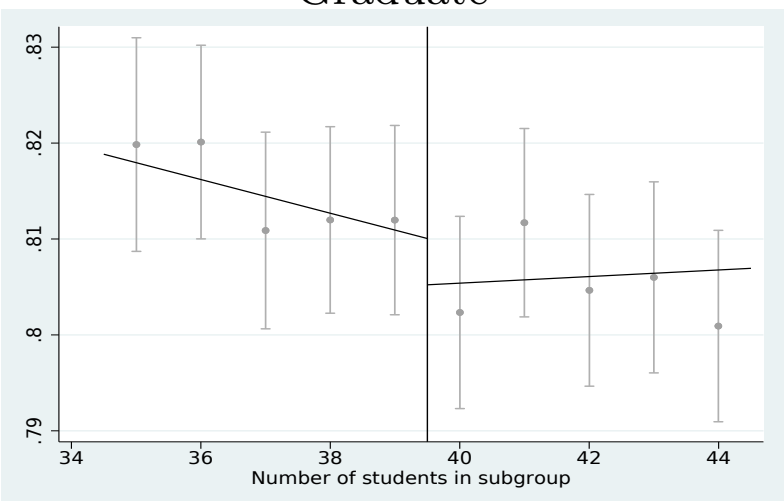
Dropout



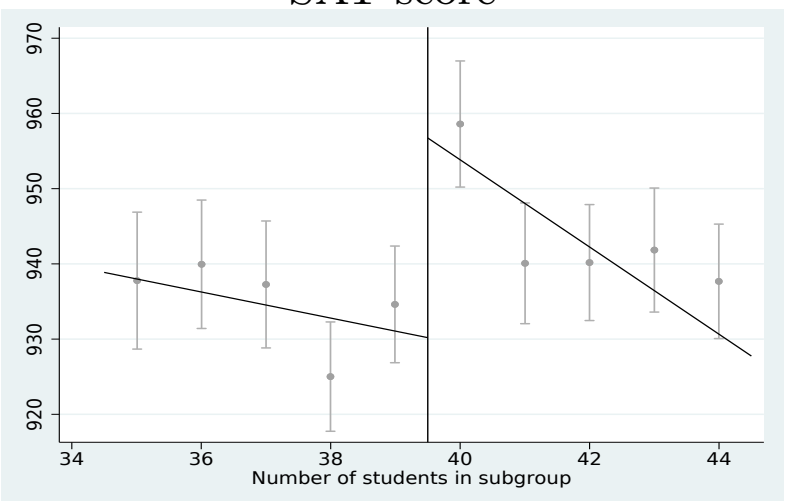
GPA



Graduate

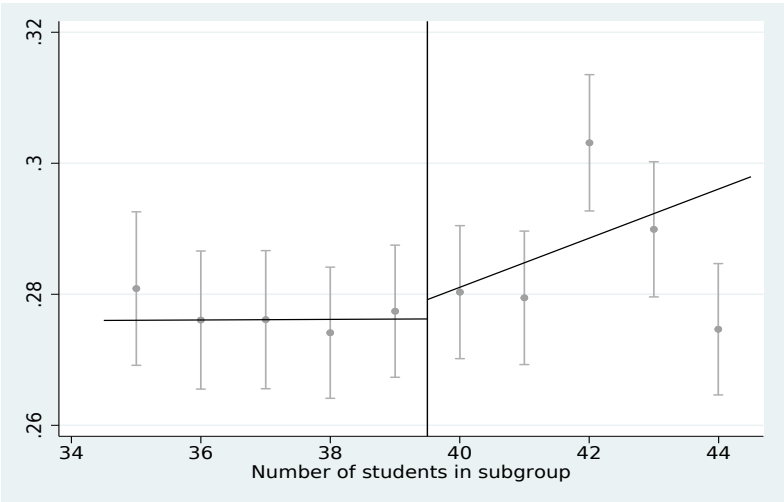


SAT score

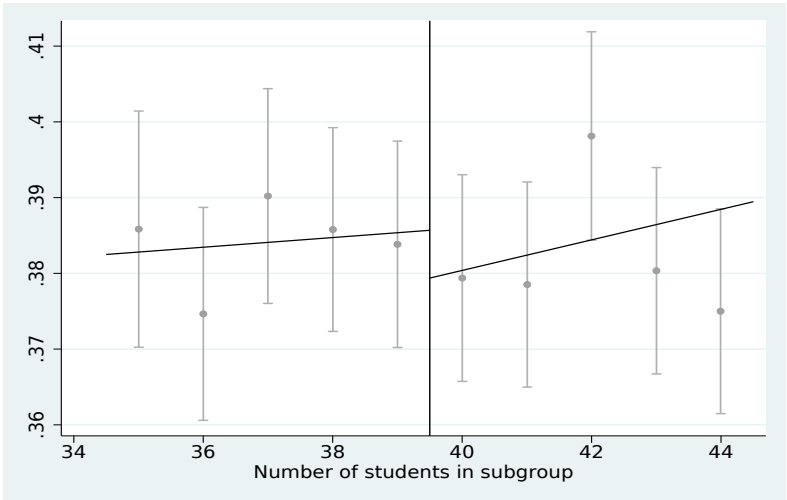


Subgroup RDD - Long-Run Plots

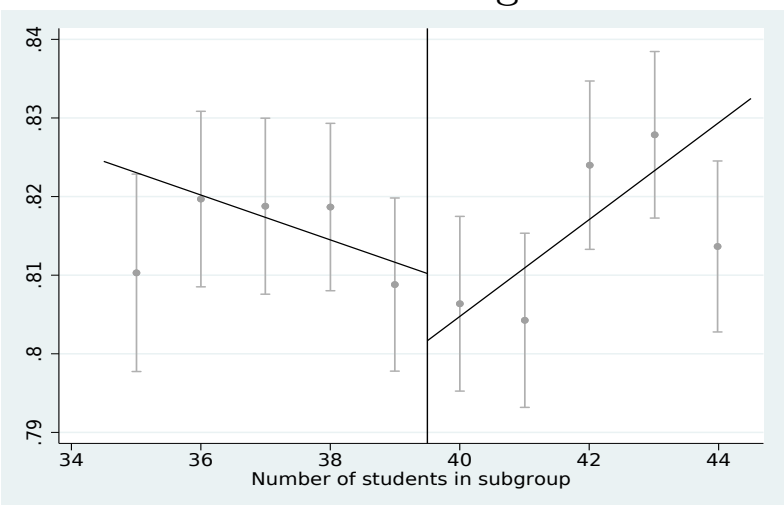
Take SAT



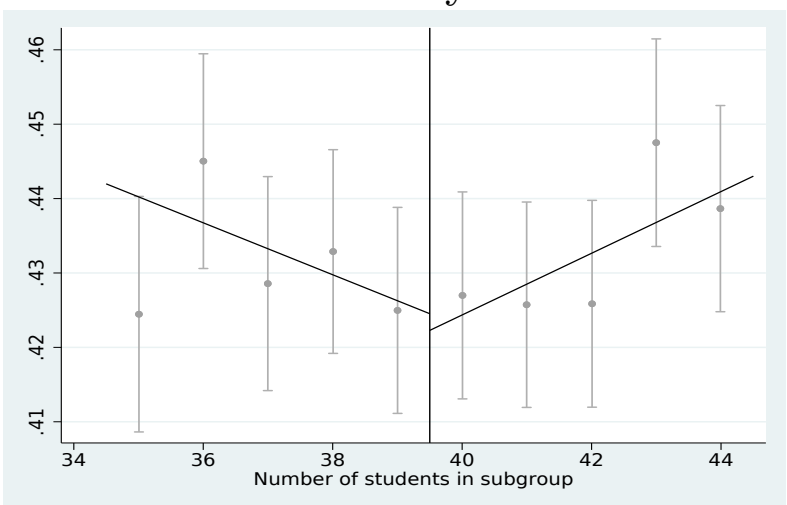
Intend 4-year



Intend College



Intend 2-year



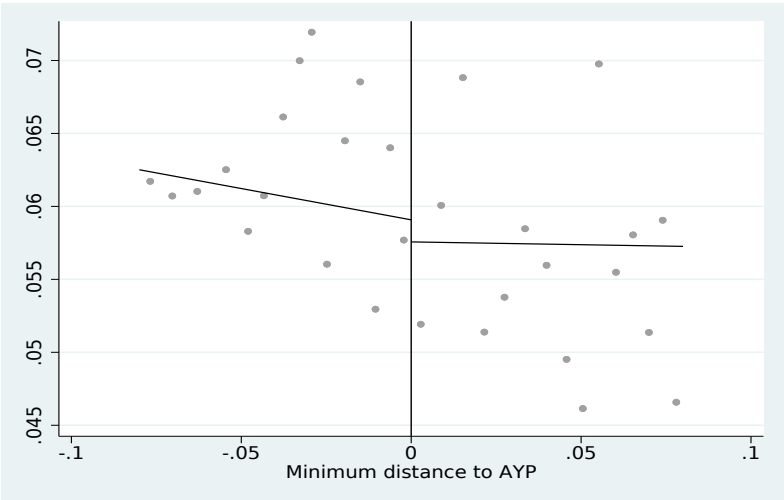
School RDD - Long-Run Outcomes

	Dropout		Graduate		HS GPA		SAT score	
Treated	0.00369	0.00454	-0.0111	-0.0134*	0.0043	-0.02	30.59**	20.69*
	(0.00388)	(0.00377)	(0.00787)	(0.00776)	(0.0271)	(0.0297)	(13.38)	(10.8)
Mean Dep. Var.	0.049		0.833		2.788		993.9	
SD Dep. Var.	0.217		0.373		0.729		192.4	
Observations	196,179	199,261	202,922	202,922	136,637	123,121	44,458	41,999

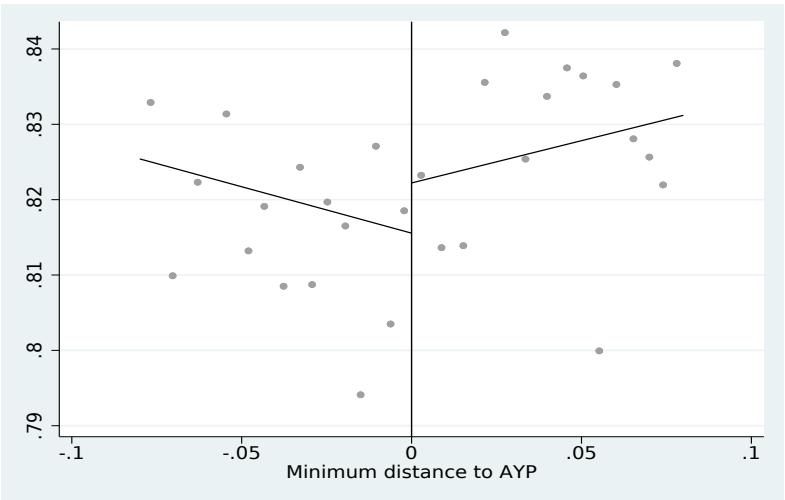
	SAT-taking		Intend college		Intend 4-year		Intend 2-year	
Treated	0.00267	-0.0083	0.00491	0.00167	0.0490**	0.0374*	-0.0428*	-0.0342*
	(0.017)	(0.0137)	(0.00894)	(0.00844)	(0.0248)	(0.0211)	(0.0219)	(0.0198)
Mean Dep. Var.	0.345		0.852		0.453		0.399	
SD Dep. Var.	0.475		0.355		0.498		0.490	
Observations	183,451	213,279	164,411	164,891	87,595	87,797	93,583	93,583
Student controls	NO	YES	NO	YES	NO	YES	NO	YES

School RDD - Long-Run Plots

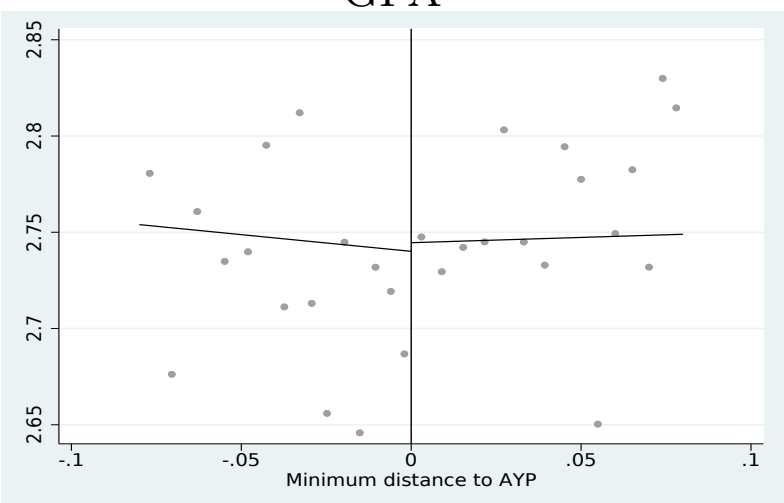
Dropout



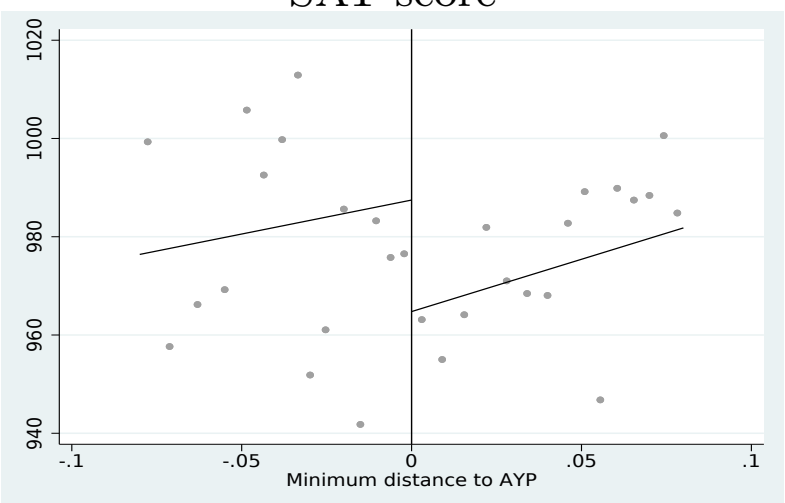
Graduate



GPA

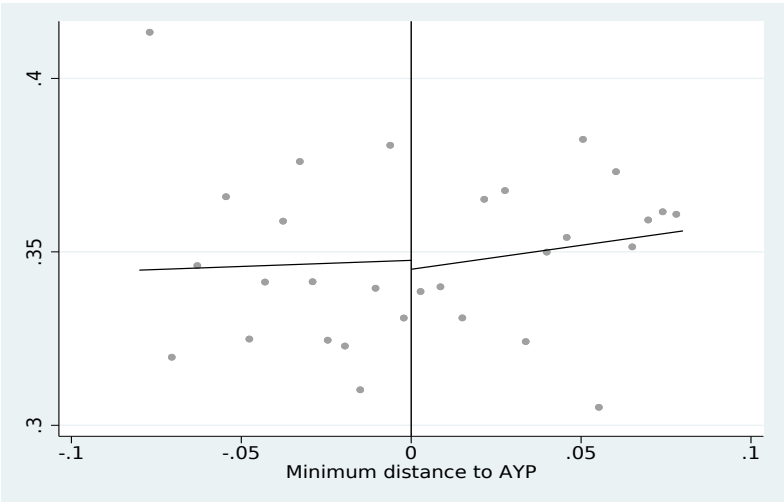


SAT score

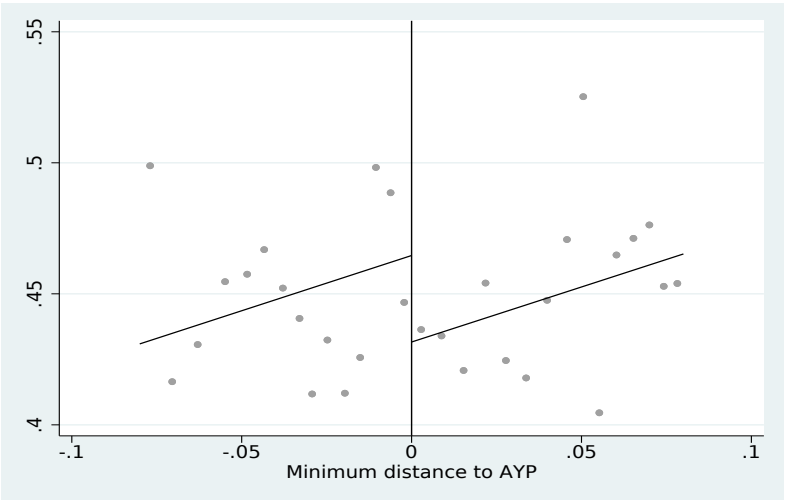


School RDD - Long-Run Plots

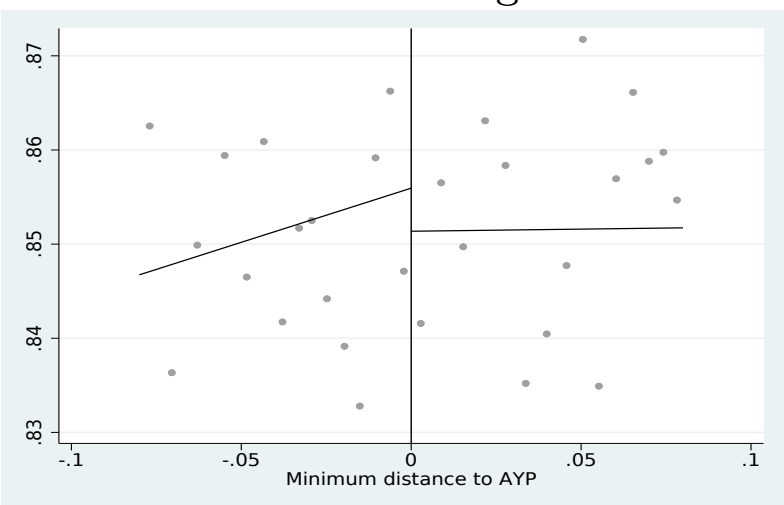
Take SAT



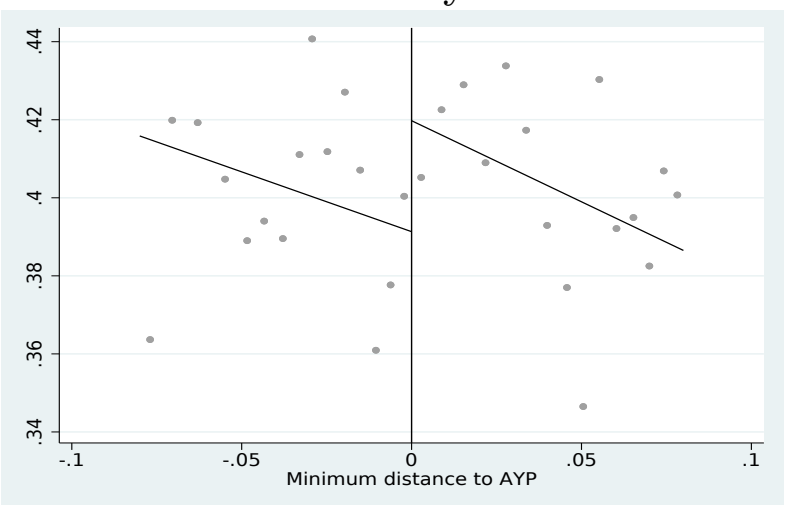
Intend 4-year



Intend College



Intend 2-year



Robustness

- Bunching / manipulation ▸ Bunching
  p-values of 0.83 and 0.42
- Balance of covariates ▸ Balance tests
- Bandwidth sensitivity ▸ Bandwidth sensitivity
- Attrition for long-run data ▸ Attrtition

What Skills Were Improved?

- One reason for the effect found on SAT scores might be a persisting increase in skills captured by test scores
  - If this is the case, accountability pressure in one year should result in higher test scores in subsequent years
- Is it specifically math and reading skills?
  - If increase in skills only pertains to math and reading, high school GPA should increase more for classes within these subjects

Subject-Specific High School GPA

Subgroup RDD

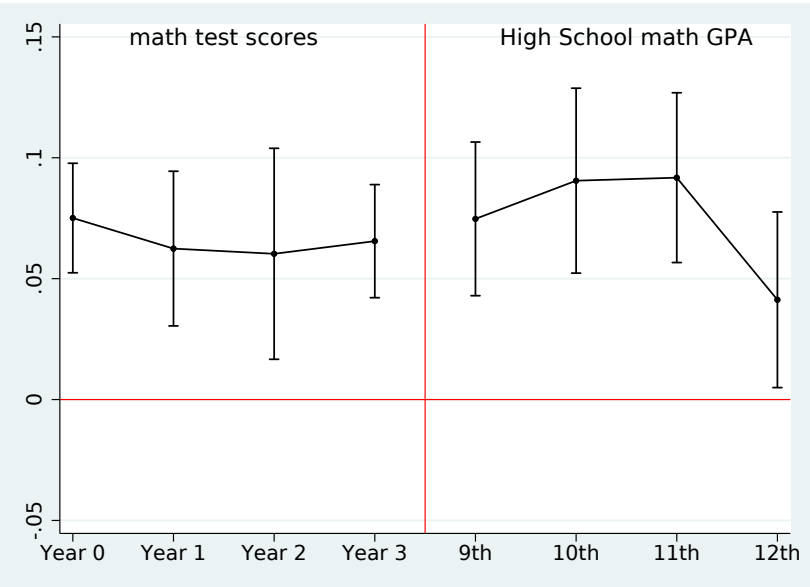
	Math GPA		Reading GPA		Non Math/Reading GPA	
Treated	0.0798***	0.0754***	0.0449***	0.0419***	0.0461***	0.0419***
	(0.0179)	(0.0175)	(0.0152)	(0.0144)	(0.0126)	(0.0116)
Observations	32,046	32,046	32,070	32,070	32,273	32,273

School RDD

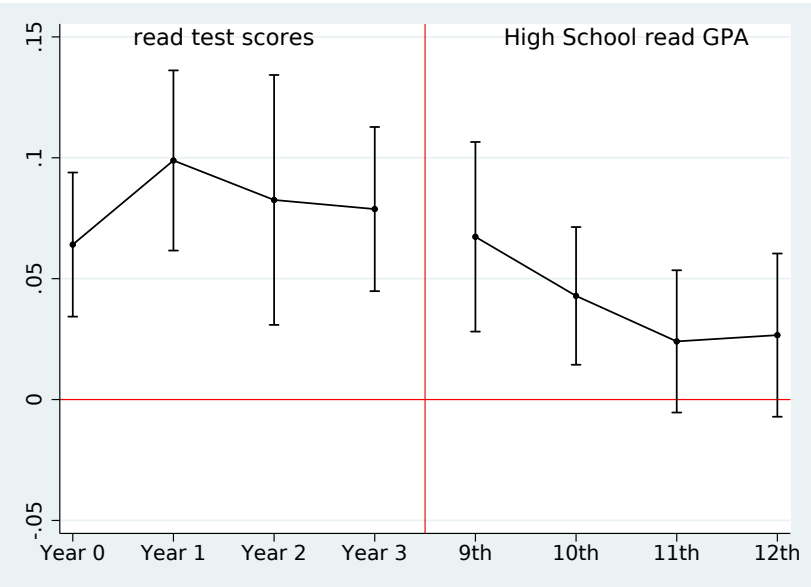
	Math GPA		Reading GPA		Non Math/Reading GPA	
Treated	0.00114	-0.0127	-0.0288	-0.0437	-0.0228	-0.0337
	(0.0586)	(0.0587)	(0.0524)	(0.0529)	(0.0435)	(0.0439)
Observations	153,865	160,147	159,987	163,260	158,686	163,620
Student controls	NO	YES	NO	YES	NO	YES

Effects on Future Academic Performance - Subgroup RD

Math

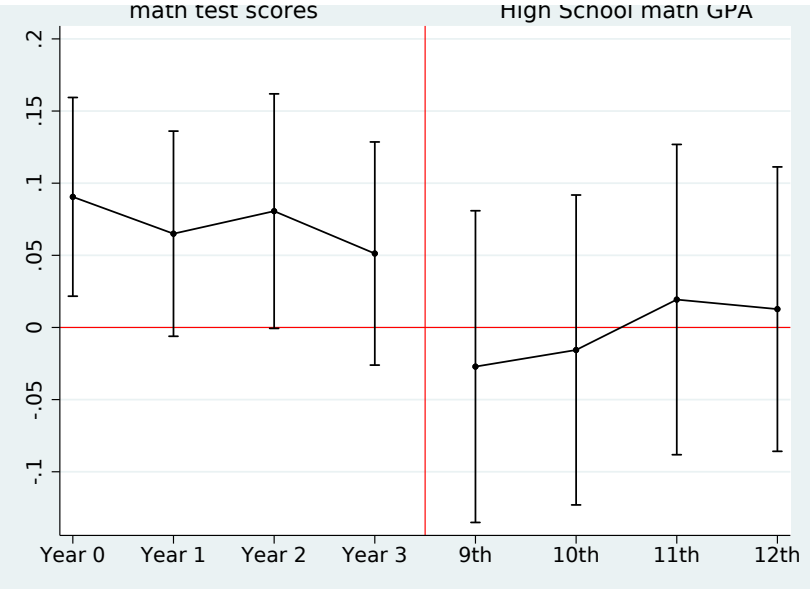


Reading

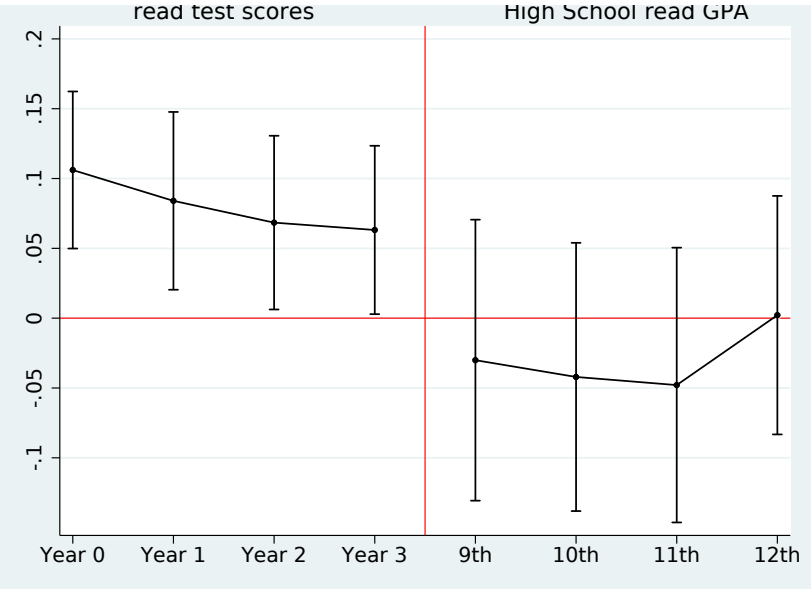


Effects on Future Academic Performance - School RD

Math



Reading



Relationship between Test Scores and Long-run Outcomes

- Long-run effects too small to detect?
- **Commensurate long-run effect**: expected LR effect given test score effect and typical relationship between test scores and LR outcomes

1. Test score effects of accountability
2. LR outcome improvement associated with higher test scores

$\rightarrow$ Commensurate effect = $\mathbf{1} \times \mathbf{2}$

- Regress: $LR_i = \beta_1 Math_i + \beta_2 Read_i + \beta_3 Controls_i + \epsilon_i$
- Test score effects: τ_{math}, τ_{read}
- Commensurate LR effect = $\hat{\beta}_1 \times \tau_{math} + \hat{\beta}_2 \times \tau_{read}$

Relationship between Test Scores and Long-run Outcomes

Not controlling for prior-year test score

	SAT_taker	SAT score	GPA	dropout	graduate	intend_4yr	intend_2yr	intend_col
math	0.102***	90.09***	0.254***	-0.0187***	0.0454***	0.136***	-0.0983***	0.0375***
	(0.000465)	(0.227)	(0.000812)	(0.000260)	(0.000440)	(0.000593)	(0.000628)	(0.000455)
read	0.0516***	70.98***	0.140***	-0.0136***	0.0256***	0.0928***	-0.0635***	0.0293***
	(0.000466)	(0.235)	(0.000821)	(0.000261)	(0.000441)	(0.000599)	(0.000634)	(0.000460)
Observations	2,004,023	657,542	1,149,794	1,591,885	1,591,885	1,330,660	1,330,660	1,330,660

Controlling for prior-year test score

	SAT_taker	SAT score	GPA	dropout	graduate	intend_4yr	intend_2yr	intend_col
math	0.0660***	65.04***	0.197***	-0.0153***	0.0386***	0.103***	-0.0718***	0.0308***
	(0.000536)	(0.261)	(0.000966)	(0.000309)	(0.000523)	(0.000704)	(0.000746)	(0.000542)
read	0.0291***	50.04***	0.101***	-0.0105***	0.0198***	0.0658***	-0.0428***	0.0230***
	(0.000519)	(0.257)	(0.000934)	(0.000298)	(0.000504)	(0.000683)	(0.000724)	(0.000526)
Observations	2,004,023	657,542	1,149,794	1,591,885	1,591,885	1,330,660	1,330,660	1,330,660

Commensurate vs. Actual Long-run Effects

Using relationship between test score gains and long-run outcomes

	Subgroup RD		
	Commensurate effect	Actual effect	Confidence Interval
Take SAT	0.006	0.004	(-0.010, 0.018)
SAT score	7.56	18.98	(5.75, 32.22)
GPA	0.020	0.026	(0.007, 0.045)
Dropout	-0.002	-0.002	(-0.011, 0.006)
Graduate	**0.004**	-0.005	(-0.014, **0.004**)
Intend 4-year	0.011	0.000	(-0.015, 0.015)
Intend 2-year	-0.008	-0.006	(-0.021, 0.009)
Intend college	**0.004**	-0.006	(-0.015, **0.003**)

Commensurate vs. Actual Long-run Effects

Using relationship between test score gains and long-run outcomes

	School RD		
	Commensurate effect	Actual effect	Confidence Interval
Take SAT	0.009	0.003	(-0.025, 0.031)
SAT score	10.86	30.60	(8.57, 52.6)
GPA	0.028	0.004	(-0.040, 0.049)
Dropout	-0.002	0.004	(-0.003, 0.010)
Graduate	**0.006**	-0.011	(-0.024, **0.002**)
Intend 4-year	0.016	0.049	(0.008, 0.090)
Intend 2-year	-0.011	-0.043	(-0.079, -0.007)
Intend college	0.005	0.005	(-0.010, 0.020)

Commensurate vs. Actual Long-run Effects

Using relationship between test scores and long-run outcomes

	Subgroup RD			*School RD*		
	Commensurate effect	Actual effect	Confidence Interval	Commensurate effect	Actual effect	Confidence Interval
Take SAT	0.010	0.004	(-0.010, 0.018)	0.014	0.003	(-0.025, 0.031)
SAT score	10.57	18.98	(5.75, 32.22)	15.21	30.60	(8.57, 52.6)
GPA	0.026	0.026	(0.007, 0.045)	0.037	0.004	(-0.040, 0.049)
Dropout	-0.002	-0.002	(-0.011, 0.006)	-0.003	0.004	(-0.003, 0.010)
Graduate	**0.005**	-0.005	(-0.014, **0.004**)	**0.007**	-0.011	(-0.024, **0.002**)
Intend 4-year	**0.015**	0.000	(-0.015, **0.015**)	0.022	0.049	(0.008, 0.090)
Intend 2-year	-0.011	-0.006	(-0.021, 0.009)	-0.015	-0.043	(-0.079, -0.007)
Intend college	**0.004**	-0.006	(-0.015, **0.003**)	0.006	0.005	(-0.010, 0.020)

Mechanisms

- How do educators increase test scores?
  - Teachers may change how they teach
  - Principals may assign students to classrooms and teachers strategically
  - I test around the subgroup cutoff for changes in teacher value-added, teacher experience, probability a teacher just transferred to the school, and class size
  - All are statistically and economically insignificant
  - Some evidence of schools retaining better teachers in the school RDD

Other Mechanisms

Subgroup RDD

	Teacher Math VA	Teacher Reading VA	Teacher Experience	Transferred Teacher	Class Size
Treated	-0.00198	0.000147	0.0478	-0.00343	-0.0804
	(0.00662)	(0.00564)	(0.164)	(0.00756)	(0.114)
Observations	272,380	260,162	311,215	191,332	240,689

School RDD

	Teacher Math VA	Teacher Reading VA	Teacher Experience	Transferred Teacher	Class Size
Treated	0.0454**	0.0643***	0.128	-0.0251	-0.468
	(0.0224)	(0.0184)	(0.758)	(0.0261)	(0.550)
Observations	171,470	133,786	214,115	209,177	171,724
Student controls	YES	YES	YES	YES	YES

Heterogeneous Effects

- SAT-takers ▸ results
- Marginal Schools ▸ results
- Subgroups below 40 ▸ results
- Marginal Students ▸ results
- By subgroup ▸ results

Conclusion

- Evidence from 2 separate RDDs to estimate effect of NCLB accountability pressure
  - Higher elementary test scores
  - Higher SAT scores in high school
  - No effects on high school graduation or attending any college
  - Mixed evidence for GPA and intention to attend a 4-year college
- Evidence on effects of test scores in subsequent years suggest persistent increase in skills captured by tests
- Null effects on other long-run outcomes could result from small, commensurate effects, but insufficient test score effects
- Suggests NCLB improved math and reading skills in the long run, but no major effect on other long-run outcomes

Balance Tests

Return

Subgroup RD

VARIABLES	Premath	Prereading	Minority	EDS	Treated next year
Treated	0.0210	0.0481	0.0231	-0.00209	0.0335**
	(0.110)	(0.110)	(0.0734)	(0.216)	(0.0132)
Observations	56,430	56,430	56,430	56,427	56,430

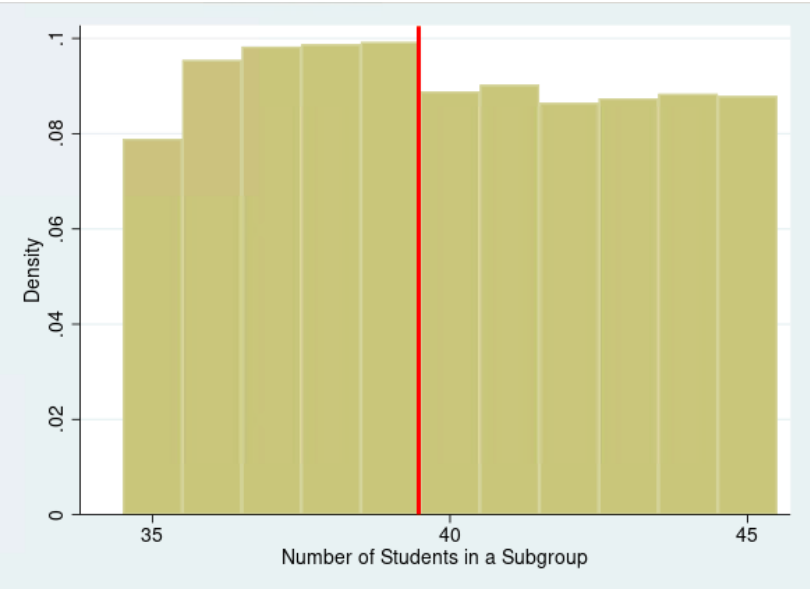
School RD

VARIABLES	Premath	Prereading	Minority	EDS
Treated	-0.0398	-0.0118	0.0711	0.0695
	(0.0656)	(0.0660)	(0.0504)	(0.0544)
Observations	450,358	450,358	450,358	450,280

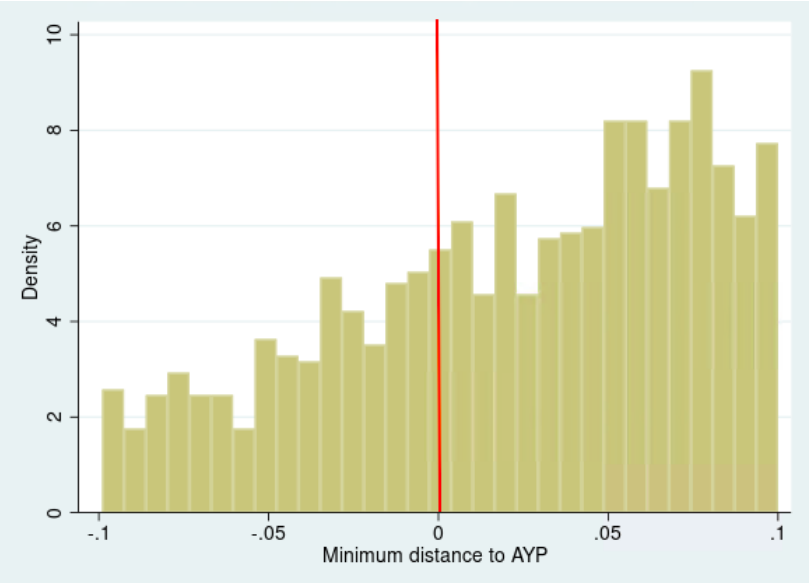
Bunching

Return

Subgroup RD



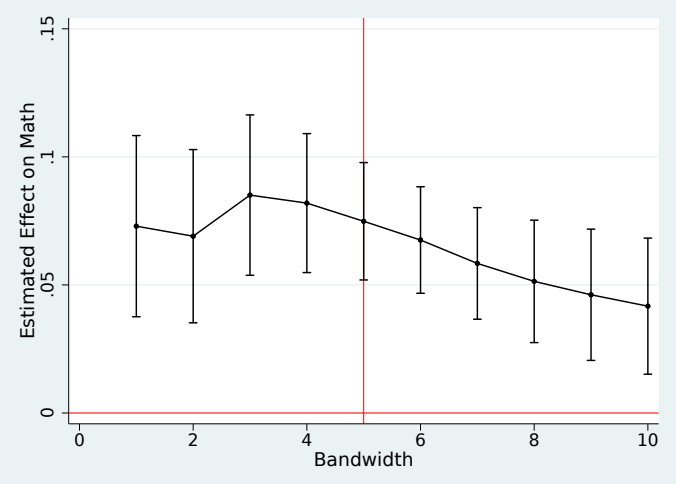
School RD

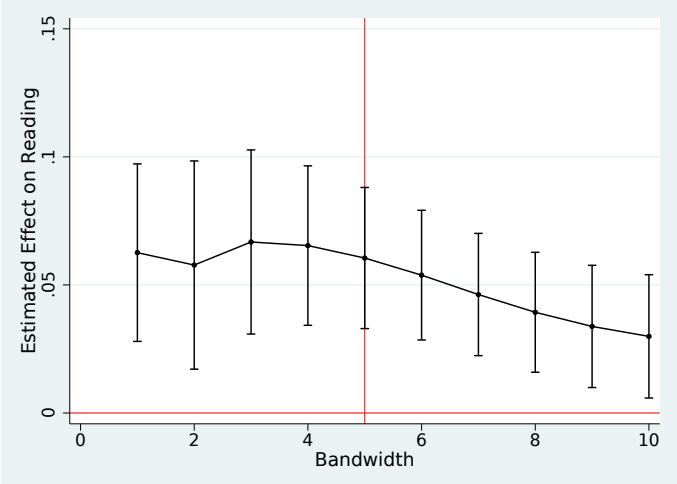


Bandwidth sensitivity - Test Effects

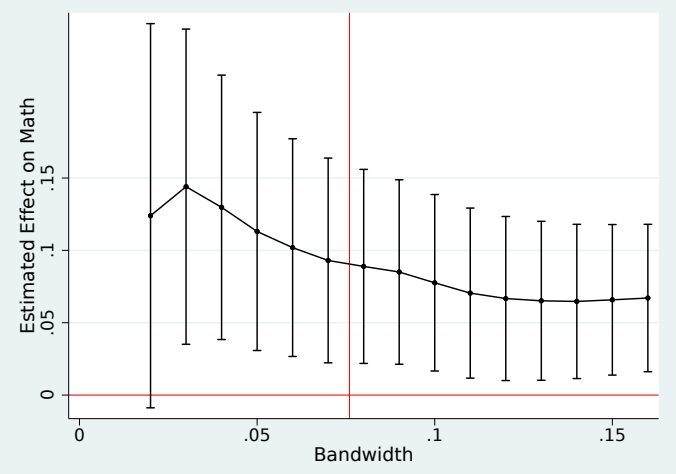
Return

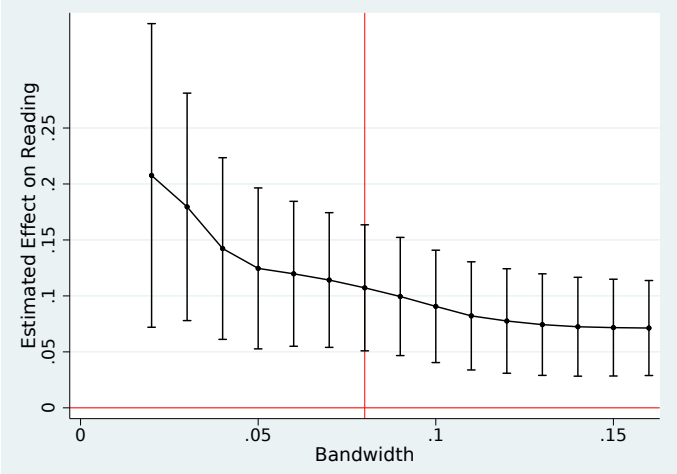
Subgroup RD





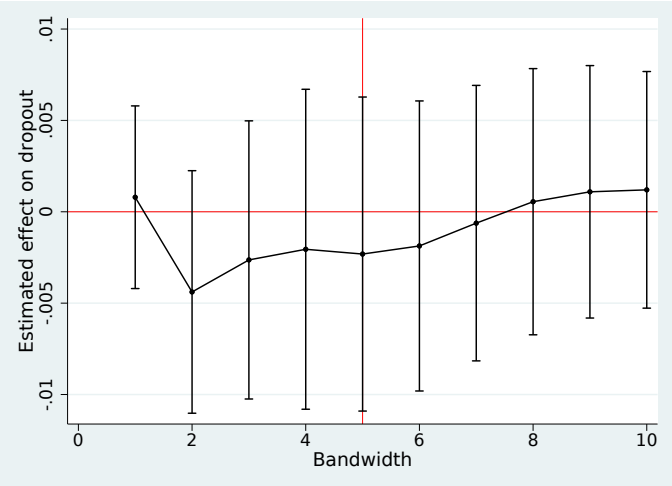
School RD

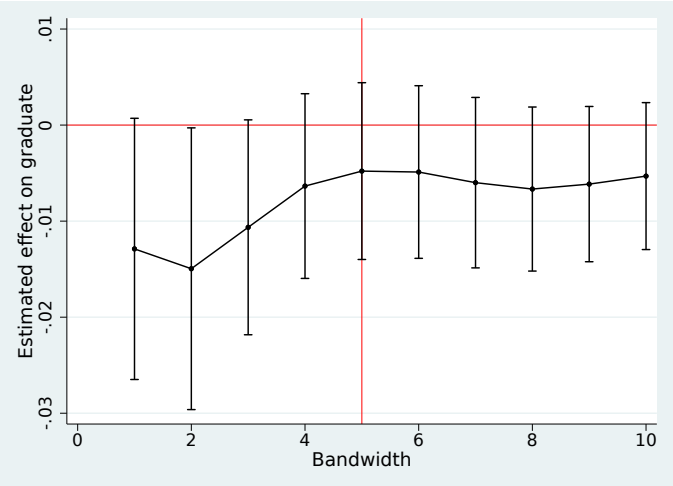


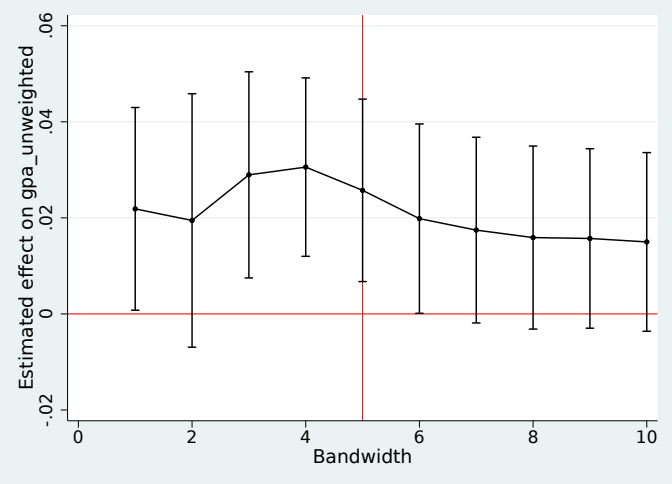


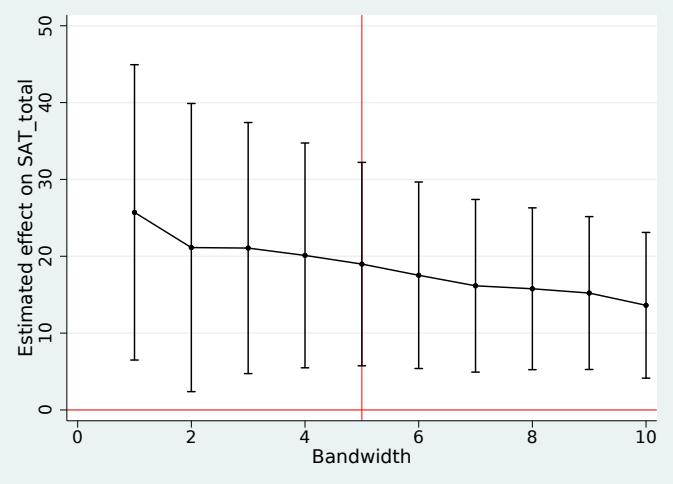
Bandwidth sensitivity - Long-run Effects

Subgroup RD



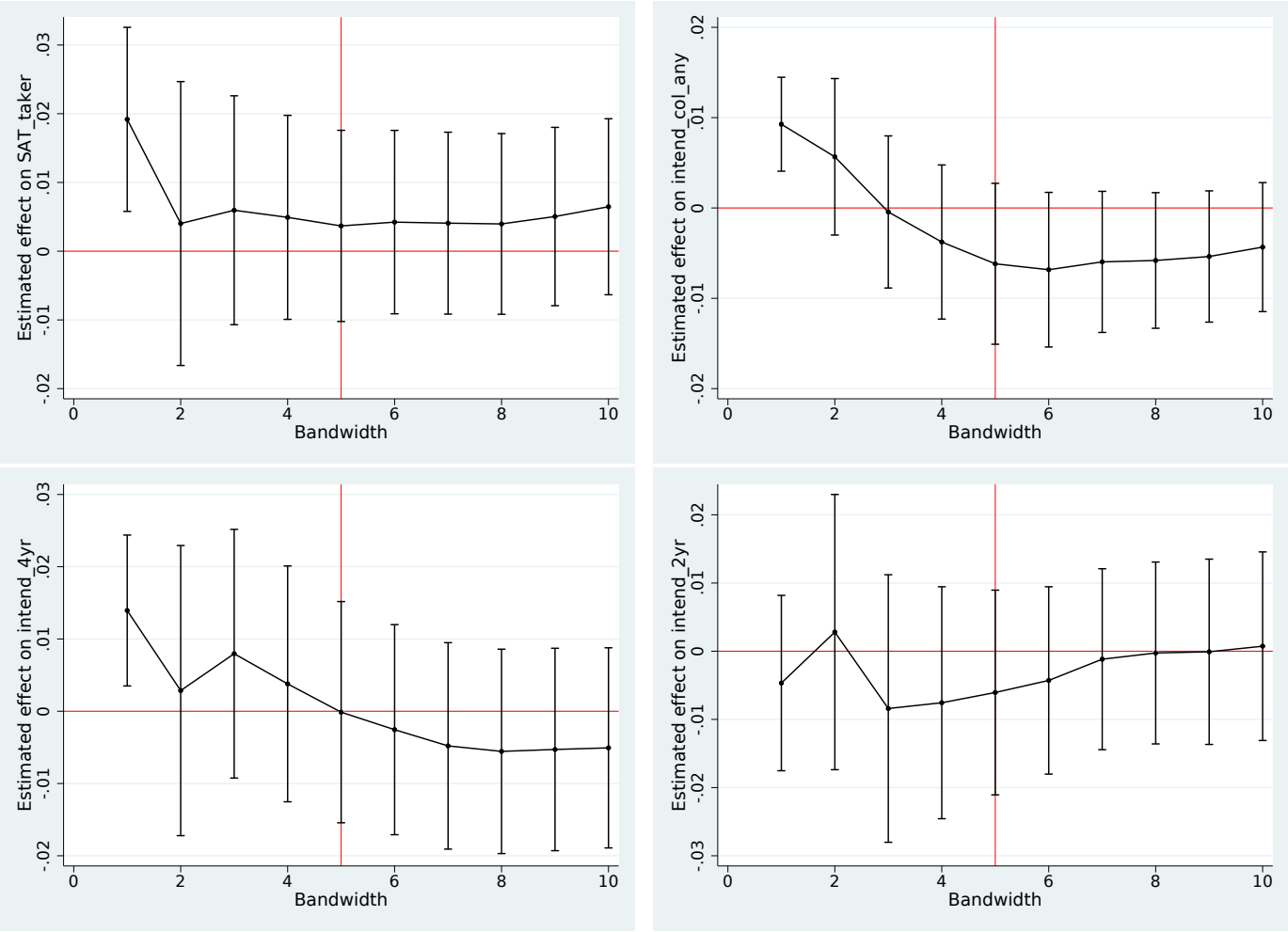






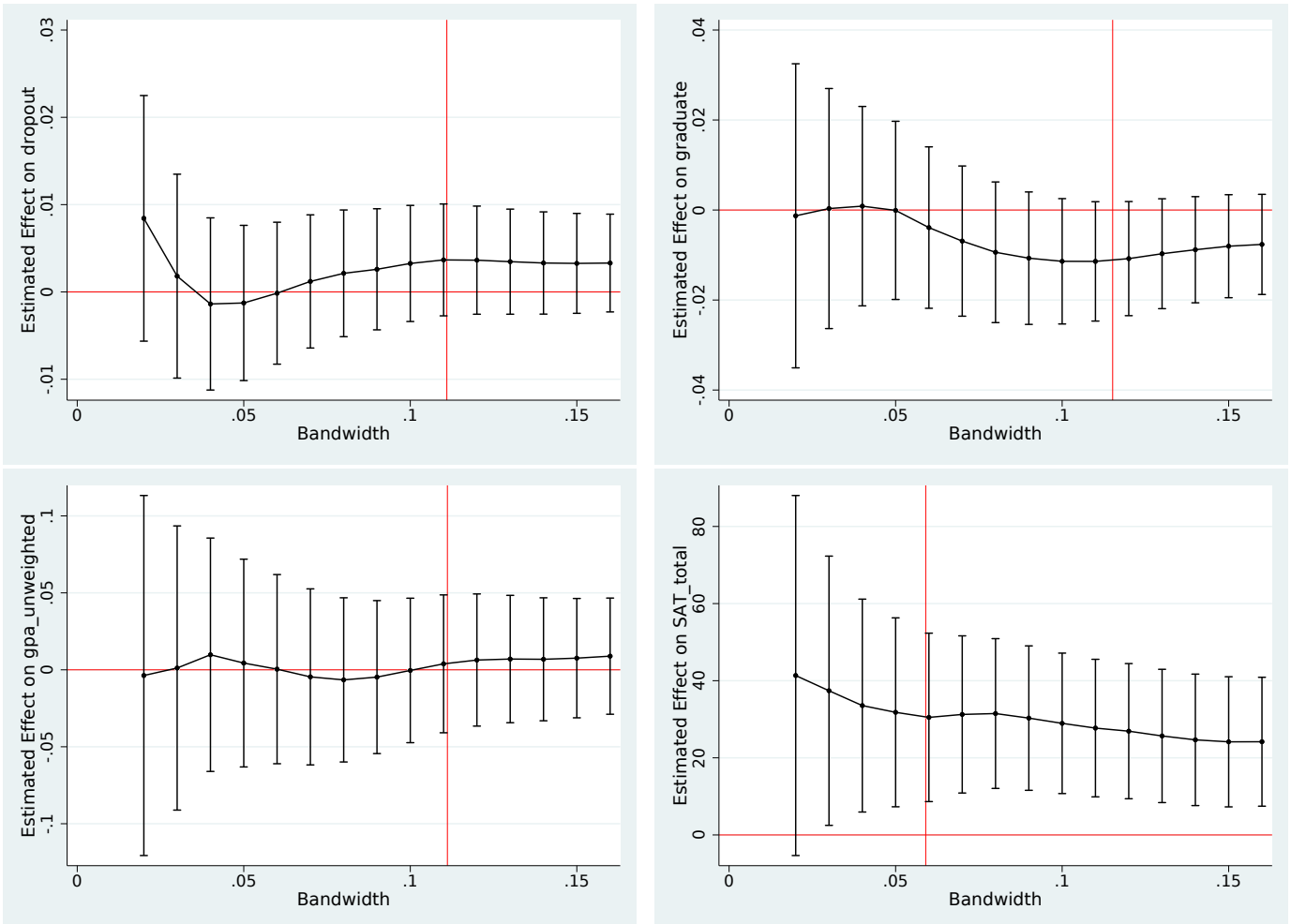
Bandwidth sensitivity - Long-run Effects

Subgroup RD



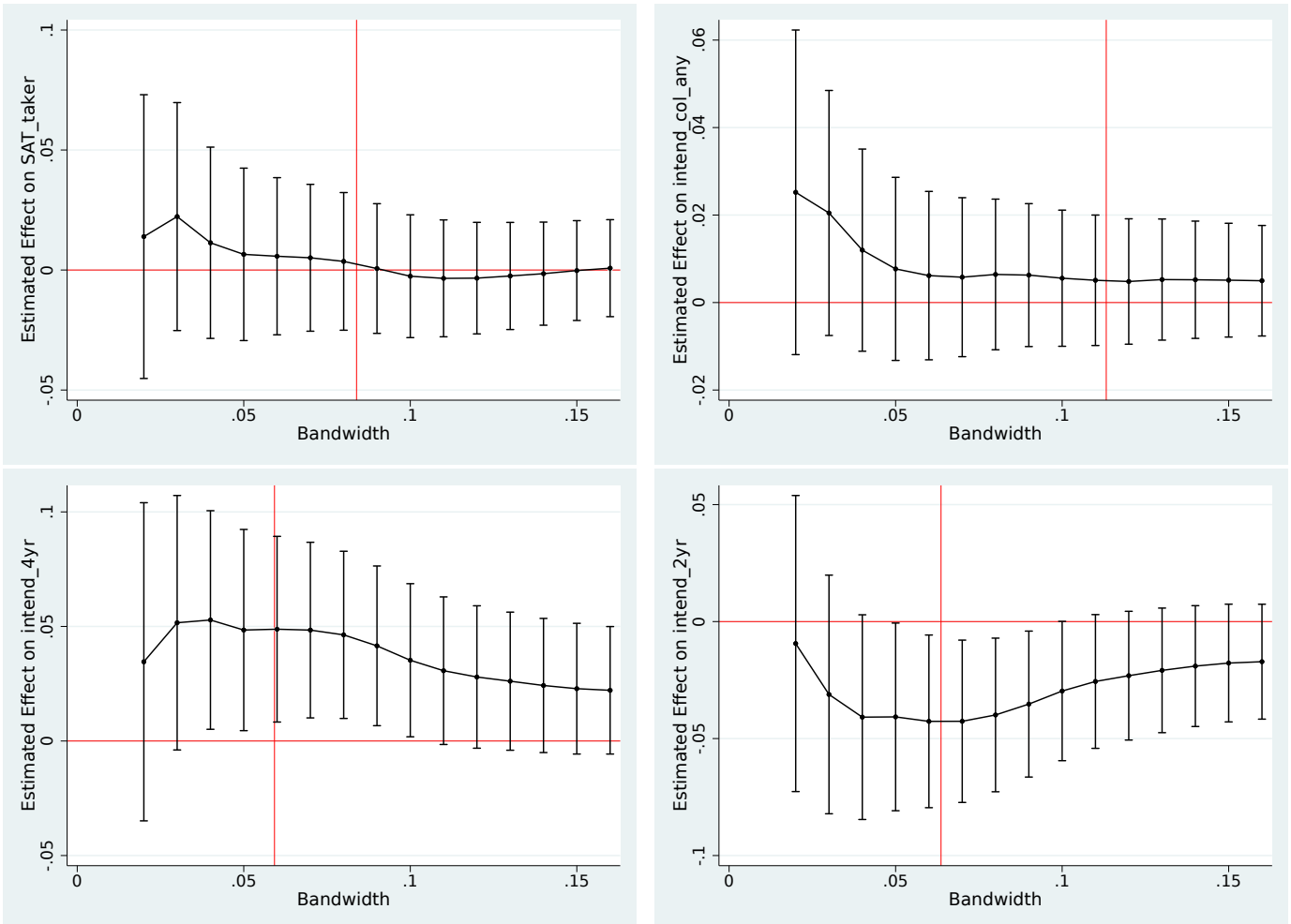
Bandwidth sensitivity - Long-run Effects

School RD



Bandwidth sensitivity - Long-run Effects

School RD



Attrition

Subgroup RDD

Missing:	SAT		GPA		Graduation		College Intention	
Treated	-0.00591	-0.00341	-0.00161	-0.00315	-0.0148***	-0.0171***	-0.00436	-0.00606
	(0.00851)	(0.00844)	(0.00932)	(0.00976)	(0.00533)	(0.00571)	(0.00678)	(0.00688)
Observations	50,747	50,747	50,747	50,747	50,747	50,747	50,747	50,747

School RDD

Missing:	SAT		GPA		Graduation		College Intention	
Treated	-0.00269	0.00829	-0.0250	-0.0197	0.00219	0.00240	0.0109	0.0130
	(0.0170)	(0.0137)	(0.0220)	(0.0216)	(0.0116)	(0.0110)	(0.0130)	(0.0126)
Observations	213,279	213,279	213,279	213,279	213,279	213,279	213,279	213,279
Student controls	NO	YES	NO	YES	NO	YES	NO	YES

Test Score Effects - SAT-takers

Return

	SAT-taker		Not SAT-taker	
Subgroup RDD				
Math effect	0.0746***	0.0748***	0.0780***	0.0509***
	(0.0263)	(0.0107)	(0.0191)	(0.0138)
Reading effect	0.0649**	0.0493***	0.0512***	0.0295***
	(0.0295)	(0.0151)	(0.0193)	(0.0109)
Observations	14,291	14,291	36,411	36,411
School RDD				
Math effect	0.0757	0.0252	0.0900**	0.0616**
	(0.0462)	(0.0266)	(0.0386)	(0.0266)
Reading effect	0.100***	0.0309*	0.0890***	0.0496***
	(0.0320)	(0.0181)	(0.0318)	(0.0185)
Observations	50,874	65,000	121,695	132,181
Student controls	NO	YES	NO	YES

Long-Run Effects - SAT-takers - Subgroup RDD

	SAT-taker		Not SAT-taker	
Dropout	0.000171	0.000161	-0.00174	-0.00162
	(0.000415)	(0.000411)	(0.00722)	(0.00694)
Graduate	-0.00627	-0.00617	-0.00659	-0.00527
	(0.00393)	(0.00383)	(0.00682)	(0.00664)
GPA	0.0426***	0.0387***	0.0139	0.0108
	(0.0136)	(0.0122)	(0.0141)	(0.0125)
Intend college	-0.00357	-0.00279	-0.00854	-0.00863
	(0.00499)	(0.00500)	(0.0109)	(0.0110)
Intend 4-year	-0.00168	-0.00429	0.00446	0.00128
	(0.0123)	(0.0115)	(0.00755)	(0.00664)
Intend 2-year	-0.00189	0.00149	-0.0130	-0.00992
	(0.0117)	(0.0105)	(0.0102)	(0.00988)
Student controls	NO	YES	NO	YES

Long-Run Effects - SAT-takers - School RDD

	SAT-taker		Not SAT-taker	
Dropout	0.000241	0.000270	0.00348	0.00316
	(0.000612)	(0.000599)	(0.00624)	(0.00617)
Graduate	-0.00250	-0.00143	-0.0118	-0.0119
	(0.00337)	(0.00341)	(0.0134)	(0.0131)
GPA	-0.00200	-0.0319	-0.00649	-0.0193
	(0.0274)	(0.0321)	(0.0363)	(0.0374)
Intend college	-0.00388	-0.00447	0.0149	0.0113
	(0.00518)	(0.00514)	(0.0144)	(0.0142)
Intend 4-year	0.0349*	0.0245	0.0453**	0.0393**
	(0.0187)	(0.0177)	(0.0196)	(0.0182)
Intend 2-year	-0.0418**	-0.0312*	-0.0306	-0.0293
	(0.0173)	(0.0163)	(0.0222)	(0.0215)
Student controls	NO	YES	NO	YES

Test Score Effects - Marginal Schools

Return

	Marginal schools		Not marginal schools	
Subgroup RDD				
Math effect	0.0518**	0.0400*	0.0930***	0.0733***
	(0.0249)	(0.0238)	(0.0250)	(0.0145)
Reading effect	0.0824***	0.0585***	0.0340	0.0204
	(0.0255)	(0.0159)	(0.0299)	(0.0138)
Observations	21,855	21,855	28,847	28,847
Student controls	NO	YES	NO	YES

Long-Run Effects - Marginal Schools - Subgroup RDD

	Marginal schools		Not marginal schools	
Dropout	0.00113	0.00183	-0.00461	-0.00484
	(0.00543)	(0.00525)	(0.00534)	(0.00545)
Graduate	-0.00852	-0.00915	-0.00233	-0.00156
	(0.00727)	(0.00715)	(0.00813)	(0.00812)
GPA	0.0239	0.0139	0.0342*	0.0300
	(0.0146)	(0.0157)	(0.0198)	(0.0196)
SAT score	15.99**	15.40***	21.20	20.43**
	(7.633)	(5.204)	(14.04)	(9.946)
Take SAT	0.0134	0.0115	-0.00135	-0.00410
	(0.00935)	(0.00953)	(0.0153)	(0.0150)
Intend college	-0.0134	-0.0137	0.000367	-0.000946
	(0.00967)	(0.00892)	(0.0103)	(0.0109)
Intend 4-year	0.0249*	0.0188	-0.0191	-0.0229
	(0.0128)	(0.0134)	(0.0178)	(0.0147)
Intend 2-year	-0.0383***	-0.0324***	0.0195	0.0219
	(0.0126)	(0.0125)	(0.0176)	(0.0148)
Student controls	NO	YES	NO	YES

Test Score Effects - Subgroups below 40

Return

	Subgroup below 40		Subgroup 40 or more	
School RDD				
Math effect	0.0640	0.108**	0.0935**	0.0416
	(0.0531)	(0.0421)	(0.0446)	(0.0253)
Reading effect	0.0591	0.0906***	0.117***	0.0378**
	(0.0411)	(0.0260)	(0.0367)	(0.0174)
Observations	29,837	23,879	137,339	173,582
Student controls	NO	YES	NO	YES

Long-Run Effects - Subgroups below 40 - School RDD

	Subgroup below 40		Subgroup 40 or more	
Dropout	0.00297	0.00291	0.00332	0.00437
	(0.00954)	(0.00958)	(0.00390)	(0.00378)
Graduate	0.00284	0.00308	-0.0126	-0.0151*
	(0.0165)	(0.0167)	(0.00803)	(0.00773)
GPA	-0.00381	-0.00456	0.00945	-0.0187
	(0.0416)	(0.0388)	(0.0282)	(0.0313)
SAT score	31.65	22.02	30.25**	20.43*
	(20.17)	(14.91)	(12.86)	(10.83)
Take SAT	-0.00487	0.00123	0.000904	-0.0102
	(0.0193)	(0.0191)	(0.0172)	(0.0140)
Intend college	0.0193	0.0196	0.00337	-0.000123
	(0.0187)	(0.0188)	(0.00918)	(0.00865)
Intend 4-year	0.0108	0.0103	0.0544**	0.0428*
	(0.0271)	(0.0243)	(0.0275)	(0.0239)
Intend 2-year	0.0117	0.0123	-0.0497**	-0.0408*
	(0.0299)	(0.0282)	(0.0239)	(0.0219)
Student controls	NO	YES	NO	YES

Test Score Effects - Marginal Students

Return

	Marginal students		Not marginal students	
Subgroup RDD				
Math effect	0.0527***	0.0531***	0.0513**	0.0117
	(0.0189)	(0.0134)	(0.0261)	(0.0116)
Reading effect	0.0321**	0.0471***	0.0168	-0.00494
	(0.0137)	(0.0101)	(0.0224)	(0.0110)
Observations	23,668	23,668	15,439	15,439
School RDD				
Math effect	0.00309	0.000655	0.0581	0.0120
	(0.0325)	(0.0291)	(0.0490)	(0.0278)
Reading effect	-0.00780	0.00729	0.121***	0.0715***
	(0.0293)	(0.0235)	(0.0327)	(0.0196)
Observations	86,279	78,854	85,425	82,221
Student controls	NO	YES	NO	YES

Long-Run Effects - Marginal Students - Subgroup RDD

	Marginal students		Not marginal students	
Dropout	-0.00454	-0.00591	-0.000570	0.000832
	(0.00443)	(0.00459)	(0.00656)	(0.00620)
Graduate	-0.00546	-0.00438	-0.0113	-0.0142
	(0.00660)	(0.00649)	(0.00893)	(0.00880)
GPA	-0.00136	-0.000319	0.0353*	0.0244
	(0.0155)	(0.0134)	(0.0209)	(0.0192)
SAT score	5.811	7.143	26.28*	16.00*
	(4.549)	(5.418)	(13.93)	(9.697)
Take SAT	0.0117	0.0121	-0.0179	-0.0228
	(0.00850)	(0.00843)	(0.0164)	(0.0155)
Intend college	-0.0151	-0.0157	-0.0231***	-0.0230***
	(0.0110)	(0.0111)	(0.00530)	(0.00610)
Intend 4-year	-0.0174**	-0.0170**	0.000976	-0.00873
	(0.00870)	(0.00769)	(0.0186)	(0.0170)
Intend 2-year	0.00231	0.00121	-0.0241	-0.0143
	(0.0156)	(0.0145)	(0.0162)	(0.0150)
Student controls	NO	YES	NO	YES

Long-Run Effects - Marginal Students - School RDD

	Marginal students		Not marginal students	
Dropout	0.00185	-0.000821	0.00497	0.00767
	(0.00655)	(0.00655)	(0.00611)	(0.00606)
Graduate	-0.00448	-0.00246	-0.0230**	-0.0276***
	(0.0130)	(0.0130)	(0.0107)	(0.0103)
GPA	-0.0384	-0.0411	0.0345	0.0107
	(0.0354)	(0.0346)	(0.0376)	(0.0371)
SAT score	19.67	19.92*	34.67**	23.02**
	(12.26)	(10.71)	(14.73)	(10.99)
Take SAT	-0.00848	-0.00892	0.0121	0.00612
	(0.0156)	(0.0164)	(0.0257)	(0.0234)
Intend college	-0.00002	-0.00363	0.0109	0.00630
	(0.0143)	(0.0140)	(0.0118)	(0.0110)
Intend 4-year	0.0222	0.0207	0.0669*	0.0504
	(0.0233)	(0.0225)	(0.0375)	(0.0327)
Intend 2-year	-0.0316	-0.0341	-0.0463	-0.0329
	(0.0249)	(0.0245)	(0.0315)	(0.0282)
Student controls	NO	YES	NO	YES

Test Score Effects - By Subgroup

Return

	White		Black		Hispanic		Disadvantaged	
Subgroup RDD								
Math effect	0.128***	0.0929***	0.0367	0.0362*	0.0507**	0.0243*	0.0691***	0.0517***
	(0.0372)	(0.0184)	(0.0309)	(0.0198)	(0.0201)	(0.0142)	(0.00985)	(0.00724)
Reading effect	0.0829	0.0604*	0.0323	0.0336**	0.0521*	0.00718	0.0576***	0.0356***
	(0.0549)	(0.0318)	(0.0229)	(0.0157)	(0.0305)	(0.0304)	(0.00748)	(0.00446)
Observations	14,641	14,641	19,478	19,478	11,918	11,918	41,923	41,923
School RDD								
Math effect	0.163***	0.0654**	0.00450	0.0105	-0.0273	0.00888	0.0228	0.0289
	(0.0616)	(0.0314)	(0.0415)	(0.0288)	(0.0634)	(0.0472)	(0.0370)	(0.0276)
Reading effect	0.221***	0.101***	-0.0162	-0.0136	0.000164	0.0506	0.0265	0.0256
	(0.0588)	(0.0277)	(0.0312)	(0.0229)	(0.0537)	(0.0406)	(0.0277)	(0.0182)
Observations	95,104	104,322	62,557	66,669	18,005	19,168	115,107	104,823
Student controls	NO	YES	NO	YES	NO	YES	NO	YES

Long-Run Effects - By Subgroup - Subgroup RDD

	White		Black		Hispanic		Disadvantaged	
Dropout	0.00749	0.00893	-0.00978***	-0.0105***	-0.00935	-0.00861	-0.00383	-0.00374
	(0.00624)	(0.00551)	(0.00239)	(0.00286)	(0.00969)	(0.00972)	(0.00524)	(0.00528)
Graduate	-0.0139	-0.0154	0.0151	0.0181**	-0.0121*	-0.0146**	0.00393	0.00410
	(0.0129)	(0.0126)	(0.00937)	(0.00853)	(0.00679)	(0.00656)	(0.00469)	(0.00473)
GPA	0.0705*	0.0679*	0.0449*	0.0432	-0.0419***	-0.0535***	0.0393***	0.0323***
	(0.0424)	(0.0367)	(0.0258)	(0.0282)	(0.00931)	(0.00853)	(0.00435)	(0.00556)
SAT score	10.79	-1.833	0.785	6.620*	28.71***	23.10***	14.78***	16.31***
	(23.63)	(15.12)	(3.116)	(3.837)	(5.745)	(6.256)	(3.400)	(3.353)
Take SAT	0.0307**	0.0269**	-0.0187*	-0.0172	0.0356***	0.0292***	0.0130**	0.0103*
	(0.0136)	(0.0116)	(0.0106)	(0.0113)	(0.00822)	(0.00882)	(0.00560)	(0.00599)
Intend college	0.0319***	0.0292***	-0.0523***	-0.0534***	0.0330*	0.0340*	-0.00309	-0.00393
	(0.00798)	(0.00954)	(0.00582)	(0.00596)	(0.0170)	(0.0179)	(0.00599)	(0.00581)
Intend 4-year	0.00743	-0.00152	-0.00791	-0.00492	0.00441	-0.00656	0.00203	-0.00275
	(0.0259)	(0.0192)	(0.0112)	(0.0123)	(0.0152)	(0.0122)	(0.00585)	(0.00526)
Intend 2-year	0.0245	0.0307	-0.0444***	-0.0484***	0.0286	0.0405*	-0.00512	-0.00117
	(0.0242)	(0.0192)	(0.0100)	(0.0114)	(0.0219)	(0.0221)	(0.00603)	(0.00547)
Student controls	NO	YES	NO	YES	NO	YES	NO	YES

Long-Run Effects - By Subgroup - School RDD

	White		Black		Hispanic		Disadvantaged	
Dropout	-0.000406	0.00288	0.00914	0.00473	0.00471	0.00119	0.00814	0.00617
	(0.00470)	(0.00475)	(0.00737)	(0.00790)	(0.0136)	(0.0133)	(0.00588)	(0.00577)
Graduate	-0.00260	-0.00913	-0.0269	-0.0206	-0.00794	-0.00359	-0.0224**	-0.0198*
	(0.00932)	(0.00910)	(0.0181)	(0.0178)	(0.0222)	(0.0213)	(0.0114)	(0.0111)
GPA	0.0567*	0.0125	-0.0643	-0.0604	-0.103	-0.0943	-0.0520*	-0.0549*
	(0.0343)	(0.0361)	(0.0449)	(0.0455)	(0.0741)	(0.0767)	(0.0307)	(0.0326)
SAT score	52.47***	28.39**	3.253	-0.554	-10.55	-6.776	10.10	5.210
	(19.13)	(14.43)	(9.379)	(8.228)	(19.28)	(12.46)	(8.644)	(6.739)
Take SAT	0.0201	0.00131	-0.0157	-0.0136	-0.0277	-0.0212	-0.0225*	-0.0220*
	(0.0247)	(0.0193)	(0.0252)	(0.0239)	(0.0226)	(0.0229)	(0.0124)	(0.0119)
Intend college	0.0103	0.00317	-0.00207	-0.00281	-0.0164	-0.0136	0.00740	0.00728
	(0.0110)	(0.0101)	(0.0166)	(0.0160)	(0.0327)	(0.0337)	(0.0129)	(0.0129)
Intend 4-year	0.0953**	0.0659**	-0.0120	-0.0118	-0.0204	-0.0141	-0.00195	-0.00254
	(0.0371)	(0.0297)	(0.0249)	(0.0248)	(0.0292)	(0.0291)	(0.0166)	(0.0155)
Intend 2-year	-0.0821***	-0.0590**	0.00993	0.00978	0.00549	0.00255	0.00818	0.00863
	(0.0318)	(0.0264)	(0.0262)	(0.0262)	(0.0348)	(0.0344)	(0.0169)	(0.0163)
Student controls	NO	YES	NO	YES	NO	YES	NO	YES