American Economic Association

The Long-Run Impacts of Specialized Programming for High-Achieving Students

Author(s): Sarah R. Cohodes

Source: *American Economic Journal: Economic Policy* , February 2020, Vol. 12, No. 1 (February 2020), pp. 127-166

Published by: American Economic Association

Stable URL: https://www.jstor.org/stable/10.2307/26888550

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/10.2307/26888550?seq=1&cid=pdf-
reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

The Long-Run Impacts of Specialized Programming for High-Achieving Students[†]

By Sarah R. Cohodes*

I evaluate long-run academic impacts of specialized programming for high-achieving students by analyzing Advanced Work Class (AWC), an accelerated curriculum delivered in dedicated classrooms for fourth through sixth graders in Boston Public Schools. Fuzzy regression discontinuity estimates show that AWC has positive yet imprecise impacts on test scores and improves longer-term outcomes, increasing high school graduation and college enrollment. These gains are driven by black and Latino students. An analysis of mechanisms highlights the importance of staying "on track" throughout high school, with little evidence that AWC gains result from peer effects. (*JEL* H75, I21, I28, J15)

Targeted programs for high-achieving students are a common but controversial educational intervention in the United States. Advocates claim they help teachers target instruction and ensure higher ability children opportunities to reach their maximum potential (Petrilli 2011, Hess 2014). Opponents see gifted education contributing to segregation and exacerbating inequality (Roda 2015). Current evidence on the effectiveness of gifted and talented programs is, however, mixed in its conclusions and restricted to short-term outcomes.

A long-established specialized program for high-achieving students in the Boston Public Schools (BPS) provides the first opportunity to study the long-term effects of gifted and talented education for elementary and middle school students in the United States. Advanced Work Class (AWC) is BPS's accelerated program for fourth through sixth graders who score well on a third grade standardized test. Students in the AWC program receive a dedicated classroom with high-achieving

peers, advanced literacy curricula, and accelerated math in later grades. While the students who participate in AWC tend to be more advantaged than BPS students as a whole, about half of AWC students are black or Latino and two-thirds of them receive subsidized school lunch. Since admission to the program is based on a third grade test score, I estimate the effect of the program using a fuzzy regression discontinuity comparing those who scored just above and just below the admissions threshold. The timing of available data makes it possible to estimate both short-run impacts on test scores and longer-run impacts on Advanced Placement (AP) course taking and test scores, SAT taking and scores, high school graduation, and college enrollment.

AWC enrollment has positive but imprecise impacts on test scores in the short-run. AWC does, however, alter students' long-run academic trajectories. AWC students participate in a more rigorous middle and high school curriculum, with AWC increasing Algebra 1 enrollment by eighth grade and AP test taking, mostly due to enrollment in AP economics. There is a large increase in high school graduation for minority students. Perhaps most importantly, AWC boosts college enrollment rates. The program increases college enrollment by 15 percentage points overall, again with gains primarily coming from black and Latino students. This results in a 65 percent increase in college enrollment for black and Latino students, most of it at 4-year institutions. Using estimated earnings associated with colleges from Chetty et al. (2017) as a measure of college quality, AWC appears to increase in college quality by about $1,750 for all students and $8,200 for black and Latino students, though these differences are not statistically significant.[1]

How does the AWC program generate these positive effects? The AWC program changes many aspects of a student's educational setting, including peer composition, teachers, and curricula that students encounter in late elementary and early middle school, as well a potential change in mindset when students are identified as "high-achieving." Attending an AWC class boosts the average test scores of peers by over 80 percent of a standard deviation, a substantial change in peer quality. However, I find little evidence to support peer effects as an explanation for AWC gains. While AWC teachers have higher gains, the change is not large enough to account for the gains in college attendance observed here. Instead, it appears that AWC is the beginning of a chain of events that causes participants to stay on track for college throughout high school.

This paper contributes to the literature in three ways. First, I provide the first evidence on the long-run effects of specialized instruction for high-achieving students. In the United States, such programs are common, with entrance to them often determined through testing.[2] There is little well-identified research on such specialized

---

[1] Effects at the mean are not statistically significant, but an examination of the distribution of college quality shows significant differences for black and Latino students.

[2] BPS does not explicitly label AWC a "gifted and talented" program, whereas the programs studied in Bui, Craig, and Imberman (2014) and Card and Giuliano (2014, 2016) are labeled as such. It is unclear how the students compare across programs. AWC-eligible students are the top 11 to 17 percent of students in BPS, but this is equivalent to national percentile rankings of about the seventieth percentile in each subject. In the district studied by Bui, Craig, and Imberman (2014), students can meet program requirements in several ways, but one of them includes scoring above the nationally normed eightieth percentile on four subjects. About 13 percent of students are identified as gifted (my calculations from Table 1). In the district studied by Card and Giuliano, 6 percent are identified as gifted and 13 percent

programs at the elementary and middle school level, with two major exceptions that use regression discontinuities to identify program effects. Bui, Craig, and Imberman (2014) finds that attending a gifted and talented program in a large urban school district does not increase most test scores, with the exception of science, despite documenting a large change in peer characteristics. Card and Giuliano (2014) studies a different large school district and finds few test score impacts for students identified as gifted by an IQ test. However, there are large gains in math, reading, and science for students who enroll in program classrooms through an alternative mechanism (Card and Giuliano 2016). These studies are unable to follow students over enough time to include outcomes beyond test scores.[3] A related intervention is accelerated course work, which I discuss more below.

Second, I provide evidence on which classroom characteristics contribute to AWC gains, which adds to the literature on peer effects, teacher quality, and access to accelerated curricula. One of the biggest changes in AWC classrooms is the change in peer quality. Sacerdote (2011) suggests that peer effects may be more related to the presences of high-ability peers, as in AWC, than to average classroom performance. This is shown empirically in multiple contexts (Imberman, Kugler, and Sacerdote 2012; Lavy, Paserman, and Schlosser 2012; Burke and Sass 2013). However, despite large changes in peer composition at exam schools, a similar intervention as AWC at a higher grade level, there are no changes in test scores or college outcomes for the students who attend them (Abdulkadiroğlu, Angrist, and Pathak 2014; Dobbie and Fryer 2014).

Additionally, students in AWC classrooms are less likely to have a novice teacher,[4] and their teachers have higher value added (though the value added differences are not statistically significant). Chetty et al. (2011) and Chetty, Friedman, and Rockoff (2014b) show that teachers can have long-term effects on student outcomes, including the college outcomes measured here.

AWC students are also exposed to a different curriculum, with the most dramatic change coming from math acceleration. Universal acceleration of math, in particular, enrolling all students in Algebra 1 by eighth grade (a goal of the AWC program), diminishes student outcome on average (Allensworth et al. 2009; Clotfelter, Ladd, and Vigdor 2015). However, these same interventions can have beneficial effects for high-achieving students who would not have otherwise been exposed to more advanced curriculum (Clotfelter, Ladd, and Vigdor 2015). A more targeted approach to acceleration involving Algebra 1 in eighth grade for a selection of students increases college readiness (Dougherty et al. 2017). In a randomized experiment, Jackson and Makarin (2018) shows that access to and support for

---

are enrolled in gifted classrooms. Within each district, all of the programs are targeted to a similar top percentage of students, but it is not possible to directly compare students' achievement levels across the three studies.

[3] Research on magnet high schools, a related intervention for older students, shows little effect on student achievement or attainment. Abdulkadiroğlu, Angrist, and Pathak (2014) and Dobbie and Fryer (2014) find that Boston and New York City students who pass admissions cutoffs for these schools attend schools with higher achieving peers but generally do not have higher test scores or college outcomes. Studies of exam schools outside of the United States tend to find more positive results. See Clark (2010) for evidence from the United Kingdom, Jackson (2010) for Trinidad and Tobago, and Pop-Eleches and Urquiola (2013) for Romania.

[4] Novice teachers consistently have lower value added on test scores than their more experienced counterparts (Rockoff 2004; Clotfelter, Ladd, and Vigdor 2007; Harris and Sass 2011; Papay and Kraft 2015).

high-quality inquiry-based math curricular materials in middle school increased test scores by about 0.1 standard deviations.

Finally, this study adds to a small body of evidence in the economics literature on the role of positive identity development in academic outcomes. The positive identity induced by the AWC offer and enrollment (being explicitly identified as high achieving) may also influence students' long-term academic trajectories. A growing body of studies, many from social psychology, show that even minor interventions that help students see themselves as successful can be influential, especially if this positive identity counters self-reinforcing negative stereotypes (see Lavecchia, Liu, and Oreopoulos 2016 for a review). Some recent papers in economics build on this research. Bettinger et al. (2018) shows in a randomized trial that an intervention to increase students' "growth mindset" increases students' academic outcomes three weeks later, with the biggest effects for students who previously thought themselves low ability. Bursztyn and Jensen (2015) shows that being in an honors class changes social norms around participating in SAT preparation, hypothesizing that social identities reinforce academic choices. The selective nature of AWC provides confirmation of one's own ability that may be particularly important for groups facing negative stereotype threat (Steele and Aronson 1995). This may be one way that AWC helps students stay "on track" through school.

This paper differs from the other papers on specialized programs for high-achieving US students in elementary and middle schools by providing causal evidence on longer-term academic outcomes. With imprecise findings on short-term outcomes but stronger evidence of program benefits on longer-term outcomes, this study supports the idea that program evaluation must include longer-term follow-up to fully measure the potential benefits of a program. Detailed data make it possible to explore mechanisms, which suggests that peers play little role in AWC gains but that staying "on-track" throughout school helps minority students maintain the academic trajectories necessary to enroll in college.

The paper proceeds as follows. The next section details the AWC program and admissions policies. In Section II, I describe the data and sample and in Section III the fuzzy regression discontinuity empirical strategy. I report results in Section IV and discuss potential threats to validity in Section V. Section VI includes a discussion of potential channels for the AWC effect, and Section VII concludes.

I. Advanced Work Class

The AWC program has been a part of BPS since before the Judge Garrity school desegregation decision in 1974.[5] It offers an accelerated curriculum to academically advanced students. AWC teachers and schools have flexibility to develop their own AWC curriculum around some common curricular standards developed by a central AWC office that supports the program across schools.[6] All AWC programs include common elements in English/language arts (ELA) and math. In ELA, the curriculum

---

[5]The allocation of AWC was part of the school desegregation plan in Boston, and, historically, AWC and exam school seats were allocated with racial preferences, in addition to the more widely known busing policy.
[6]I thank Ailis Kiernan of the BPS AWC curriculum office for describing the program to me.

includes novels and longer texts, some from a required list, whereas typical BPS classrooms are more likely to use anthologies and excerpts. There are required writing responses to the texts, and instruction focuses on key questions that ask students to write responses to the material they have covered. In mathematics, fourth grade is used as a foundation to make sure all AWC students are at the same level, and then the math curriculum is accelerated in fifth and sixth grades, so that students cover additional material. The goal is for students to take calculus in their senior year of high school, which entails pre-algebra in seventh grade and algebra in eighth grade. There are no formal science or social studies requirements, but program instruction again uses key questions. There are also non-curricular aspects to the program. Students are in classrooms with higher achieving peers and program-specific teachers, and they have been identified as high-achieving by qualifying for the program.

Students are accepted into the program by their score on a nationally normed standardized exam offered in the fall of third grade. All third grade students are tested, with an alternative exam offered for Spanish-speaking students.[7] Acceptance to the program is based on passing a threshold that incorporates both the math and reading portions of the exam. The thresholds may change each year depending on the number of available seats and scores of the third grade. In the third grade cohorts from 2001 to 2012, the top 11 to 17 percent of the third grade test-takers are offered the program, with more students becoming eligible as additional school AWC programs were put in place.[8]

Importantly, not every BPS school that serves third graders has an AWC program. Students are guaranteed a seat in the program if they score above the cutoff but may have to switch schools. Some families choose not to accept the AWC offer if it involves a school switch. Families are notified of AWC program acceptance in the winter, and they must return a school choice form to select an AWC school. Families and teachers may appeal the AWC decision, and appeals are considered on a case by case basis. Students are typically offered a spot in AWC in fifth grade if they attended in fourth grade, though students must make academic progress in AWC. In fifth grade, all students, including those attending AWC, are retested, and sixth grade acceptance to AWC is based on the retest. In some cases, students must switch schools again to find a school that offers AWC in sixth grade. Since the BPS school choice process typically takes place prior to kindergarten and sixth grade, accepting the AWC offer also involves the affirmative process of returning a school choice form in a grade level that many families are not primed to do so. Thus, one reason for the somewhat low take-up rate of AWC is that the default option (not returning a school choice form) results in no AWC enrollment.

Figure 1 shows how the threshold works in practice. Enrollment in fourth grade AWC jumps by about 35 percentage points at the cutoff. There is less than perfect compliance with the offer of enrollment, since many families choose not to enroll, as described above. However, very few students beneath the threshold enroll

---

[7] Boston residents who do not attend BPS schools are also offered the opportunity to take the exam. There are two citywide AWC programs for Spanish-speaking students.

[8] Notably, while these are the top achievers in BPS, the nationally normed percentile rank equivalent of the threshold is around the seventieth percentile in both math and reading.
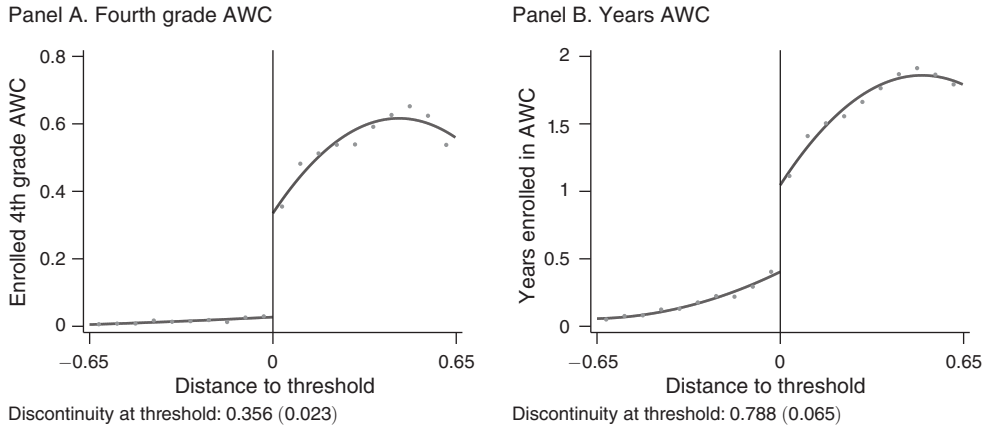
Panel A. Fourth grade AWC

Panel B. Years AWC

Discontinuity at threshold: 0.356 (0.023)

Discontinuity at threshold: 0.788 (0.065)

FIGURE 1. AWC ENROLLMENT BY DISTANCE TO ELIGIBILITY THRESHOLD

Notes: This figure shows AWC enrollment by the running variable for the third grade cohorts from 2001 to 2005 within the bandwidth of 0.65 around the eligibility threshold. A quadratic fit is imposed on either side of the threshold. Each dot represents the average enrollment for a bin of width 0.065. Panel A shows enrollment in fourth grade AWC, and panel B shows years of enrollment in fourth through sixth grades.

Source: Author's calculations from BPS and DESE data

in the program through the appeals process if they are not eligible. Panel B shows a jump in years of enrollment at the threshold of about three-quarters of a year. Students who score directly under the threshold do have an increase in enrollment in the program, up to about 0.4 years of attendance. This is mostly due to students who qualify for sixth grade AWC on their fifth grade test. As described in detail later, I employ a fuzzy regression discontinuity empirical strategy to estimate program effects which accounts for imperfect compliance to the threshold rule.

II. Data and Descriptive Statistics

A. *Data*

BPS provided records of all third grade test-takers in the fall of 2001 to the fall of 2012. The exam was the Stanford 9 for the fall 2001 to 2008 cohorts and TerraNova for fall 2009 cohorts forward, both nationally normed standardized tests with reading and math sections. BPS also provided lists of students enrolled in AWC. These records form the basis of the sample, which is then linked to data from the Massachusetts Department of Elementary and Secondary Education (DESE).

DESE provided data on student enrollment and demographics, state standardized exams, AP and SAT test taking and test scores, and National Student Clearinghouse (NSC) records of college enrollment for 2001–2015. The Student Information System (SIMS) records provide demographic characteristics and status as a special education student, English learner, or subsidized lunch recipient for third graders. It also includes attendance and school enrollment information. Third grade test scores from the Massachusetts Comprehensive Assessment System (MCAS) provide an

alternative measure of student achievement from the exam used to determine AWC eligibility.[9] Third grade ELA MCAS scores are available for all cohorts, and third grade math MCAS scores are available since 2006.[10] With the full universe of Massachusetts public school students in the data, it is possible to follow students throughout their academic careers even if they leave BPS, as long as they remain in Massachusetts public schools.

For school years 2010–2011 to 2013–2014, DESE also provided Student Course Schedule (SCS) and Education Personnel Information Management System (EPIMS) records. These data permit linking students and teachers to specific classrooms and courses, generating classroom peer characteristics and teacher characteristics, including teacher value added, for fourth through sixth grade classrooms in recent cohorts. Peer characteristics are calculated using third grade student characteristics, grouped by the course identified in the student-teacher-course link. Teacher valued-added calculations for the fourth through sixth grade math and ELA classrooms use a specification with lagged test scores, their squares and cubics, demographics, and peer demographics and test scores following Kane and Staiger (2008). I use a leave-year-out estimator to reduce bias, as indicated in Chetty, Friedman, and Rockoff (2014a, b).

Outcome measures link the records of third graders to their MCAS scores across their academic careers, AP and SAT test taking and test scores, high school graduation indicators from the SIMS database, and indicators of college enrollment from the NSC. Some outcomes are based on projected senior year in high school, determined by adding 10 to the fall year of third grade. Unless otherwise specified, all outcome data comes from DESE.

The following list summarizes the main study outcomes:

- *Enrollment*: Enrollment indicators in fourth through twelfth grade track enrollment at any BPS school, a BPS exam school (a district seventh through twelfth grade magnet school with acceptance determined by test), Boston charter schools, and non-Boston Massachusetts public schools (including non-Boston charters). On-time progress in each grade is an indicator for a student being in the appropriate grade after third grade, assuming yearly promotion, for students observed in the data.
- *MCAS*: MCAS raw scores are standardized among the full state population to be mean zero and standard deviation one in each subject and grade. In fourth through eighth and tenth grades, all students are tested in math and ELA in most years. Fourth, seventh, and tenth grades also include a writing exam, scored on two dimensions: topic development and writing composition (grammar). Science is included in fifth, eighth, and tenth grades.

---

[9] Since MCAS exams are administered in the spring after students and their families are notified of AWC eligibility, it is possible that being above the threshold for AWC acceptance has an effect on third grade MCAS scores. This would not be an effect of enrolling in the program but perhaps an independent effect on self-esteem due to knowledge that one was above the threshold. However, in practice, third grade MCAS scores are not discontinuous at the threshold, as seen in online Appendix Figure B.2.

[10] The No Child Left Behind Act (NCLB) required testing in both math and reading in grades 3 through 8 and once in high school. Prior to implementing NCLB testing requirements in the 2005–2006 school year, Massachusetts had some exams in all grades 3 through 8 and 10, but not in all subjects.

- *Exam school application*: In addition to enrollment in exam schools, this study includes exam school application and offer indicators and scores on the Independent School Entrance Exam, the optional test used for exam school admission.[11] Students must be present in the data in the relevant grade (seventh or ninth) to contribute to these outcomes.
- *AP and SAT*: AP and SAT are observed for the cohorts of third graders who are in twelfth grade in projected senior years of 2011 through 2015. I report outcomes for test taking, passing exam thresholds, and scores (1–5 for AP, 200–800 for each SAT section). Test taking and passing test threshold outcomes are generated for those present in the data in twelfth grade.
- *High school graduation*: I observe high school graduation from any school in Massachusetts for projected senior years of 2011 through 2015. I observe five-year high school graduation for one less cohort. These outcomes are unconditional for those present in the data in twelfth grade.
- *College*: NSC data is available for third graders with projected senior years of 2011 to 2015. College enrollment indicates enrollment within six months of time since expected high school graduation and is differentiated by college type (two- or four-year). All college outcomes are constructed for students who were sent to the NSC for matching.[12] College quality is measured by linking colleges to information from Chetty et al. (2017) on the estimated 2014 earnings of individuals who attended those colleges from the 1980–1982 birth cohorts. Each student is assigned the mean earnings of college attendees of whichever college they attended by gender (regardless of time of enrollment), and non-attendees are assigned the mean earnings for those who did not attend college by gender.
- *Peer and teacher characteristics*: Classroom characteristics are available for third grade cohorts from 2007 through 2012, for whom student-teacher-course links are observed. Peer characteristics include demographics, special education, English learner, subsidized lunch status, and test scores from third grade, averaged at the classroom level. Teacher characteristics include value added, years of experience, and novice status.

In order to follow a consistent sample of students throughout the paper for whom college outcomes are available, the main results follow third grade cohorts from 2001 to 2005.

B. *Descriptive Statistics*

Third graders in BPS as a whole generally come from a disadvantaged background. As shown in column 1 of Table 1, which shows student characteristics measured in third grade for study cohorts from 2001 to 2005, most third grade BPS

---

[11] The data for these outcomes are the same data used in Abdulkadiroğlu, Angrist, and Pathak (2014). I thank the authors and BPS for sharing these data.

[12] In the regression discontinuity sample, all students in the 2001 cohort were sent to NSC for matching, 90 percent of the 2002 cohort, 83 percent of the 2003 cohort, 78 percent of the 2004 cohort, and 73 percent of the 2005 cohort.

TABLE 1—SUMMARY STATISTICS

	All students (1)	Enrolled in 4th grade AWC (2)	RD sample (3)
Panel A. Demographics			
Female	0.482	0.524	0.508
Black	0.478	0.236	0.378
Latino	0.303	0.198	0.240
White	0.124	0.255	0.206
Asian	0.086	0.304	0.169
Other race	0.008	0.007	0.007
Subsidized lunch	0.841	0.669	0.764
English learner	0.188	0.171	0.137
Special education	0.196	0.012	0.054
Third grade ELA MCAS	−0.753	0.536	0.142
Panel B. AWC enrollment			
Fourth grade AWC	0.067	1.000	0.187
Fifth grade AWC	0.068	0.919	0.190
Sixth grade AWC	0.097	0.777	0.264
Years AWC	0.232	2.696	0.641
Panel C. MCAS standardized index			
Fourth grade	−0.529	0.716	0.230
Tenth grade	−0.510	0.670	0.169
Panel D. High school milestones			
Took any AP	0.357	0.774	0.557
Took SAT	0.675	0.916	0.823
On-time high school graduation	0.737	0.910	0.841
Panel E. On-time college enrollment			
Any college	0.430	0.718	0.590
Four-year college	0.324	0.668	0.508
Two-year college	0.107	0.049	0.082
College quality ($2014)	37,525	56,121	45,105
Observations	20,250	1,356	6,475

Notes: Mean values of each variable are shown by sample. Column 1 is the full sample of third graders enrolled in BPS in the fall years from 2001–2005. Column 2 restricts that sample to students enrolled in AWC in fourth grade. Column 3 restricts the full sample to those within a bandwidth of 0.65 around the eligibility threshold. College quality earnings outcomes are measured by the estimated 2014 earnings of college attendees from the 1980–1982 birth cohorts from Chetty et al. (2017). Students are assigned the earnings outcomes of the college they attend by gender even if they are not on-time attendees. Students who do not attend college are assigned the outcomes for non-attendees of the same gender.

Source: Author's calculations from BPS, DESE, and NSC data

students receive subsidized lunch (84 percent) and are nonwhite (88 percent). About 19 percent of all third graders are English learners, and 20 percent are special education participants. Third grade test scores are well below the state average. Compared to the BPS population, AWC participants are more advantaged. About 7 percent of fourth and fifth graders are enrolled in AWC, and 10 percent of sixth graders. Column 2 of Table 1 indicates that those who enroll in fourth grade AWC are more likely to be girls, less likely to be black or Latino, more likely to be white or Asian, and less likely to received subsidized lunch or be an English learner. Very few AWC enrollees are also identified as receiving special education services. They score over half a standard deviation (σ) above the state mean on third grade MCAS, and most

students who enroll in fourth grade continue on in AWC in the subsequent years. Importantly, while this population is less disadvantaged than the BPS population as a whole, 67 percent of AWC enrollees still receive subsidized lunch.

Finally, students near the threshold for AWC qualification (column 3) are generally quite similar to AWC enrollees but slightly more disadvantaged, with average third grade test scores 0.4σ lower than enrollees but still above the state mean. This makes sense, since it includes students on both sides of the eligibility threshold. The differences in racial composition between the regression discontinuity sample and students enrolled in AWC comes from two factors: the prevalence of test score by race at various achievement levels and differential take-up by race. Black and Latino students are less like to have third grade scores that put them far above the eligibility threshold, and Asian students are more likely to have high scores. Online Appendix Table B.2 describes which student characteristics are associated with AWC enrollment, both above and below the threshold, not limited to the RD sample. Asian students are the racial group most likely to enroll, if given an offer. Underneath the threshold, "always takers" are typically high-achieving students and are less likely to be special education students.

In terms of outcomes, as shown in panels C and D of Table 1, AWC students outpace their peers in BPS. For MCAS scores, Boston students typically score half a standard deviation below the state mean, whereas AWC students score 0.7σ above the mean. AWC students are much more likely to take an AP test or the SAT and to graduate high school. Finally, 72 percent of AWC students enroll in any college within 6 months of expected high school graduation, whereas 43 percent of the district as a whole does. AWC students certainly do better on important outcomes than students as a whole in BPS. But it is unknown whether this difference in outcomes is due to enrollment in the program or selection bias. It is possible that students who enroll in AWC would have done just as well in absence of the program, perhaps because they are high-achieving students or because of family support. This paper determines if any of these positive outcomes associated with AWC students can be causally attributed to the program.

III. Empirical Framework

A. *Eligibility for AWC*

All third grade students in Boston have the opportunity to qualify for AWC by taking an exam in third grade.[13] The sample includes all test-takers, including private school students and students who repeated third grade, in order to identify the AWC cutoff among the entire distribution of third grade test-takers.[14] BPS

---

[13] There is a second opportunity to enroll in AWC with a fifth grade exam for entrance to AWC in sixth grade, which also sometimes disqualifies students from the program if they do not meet the threshold a second time. Estimates from this secondary threshold are shown in online Appendix D, but should be interpreted with caution due to the dual nature of the cutoff (qualifying some additional students but disqualifying some already qualified students).

[14] This means that students can be in the sample in multiple years. In practice, this happens very rarely, as grade repeaters are not near the threshold for AWC qualification, so they are not in the RD sample. Outcome regressions restrict the sample to BPS students at baseline, which excludes a small number of students who are enrolled in

calculates AWC eligibility as follows. The third grade math and reading raw scores are standardized to be mean zero and standard deviation one, with zeroes substituted for missing scores. These math and reading *z*-scores are then averaged together, and eligibility is determined using this combined score. The particular year's cutoff is based on number of AWC seats available and the current year's test score distribution, with about the top 11 to 17 percent of students eligible in a given year. More seats are offered in more recent years. Students who take the Spanish language exam may qualify under either exam. The district offers more students seats than there are spots available, knowing that not all students take up the offer.

BPS provided the exact cutoff score for fall cohorts from 2003 onward. Unfortunately, cutoffs for the two previous cohorts were not retained. In order to determine the cutoffs in fall 2001 and 2002, I reconstruct the BPS eligibility process in the data and test each possible combined score to see how it predicts enrollment in fourth grade AWC. For the years without an official threshold available, I select as a given year's threshold the cutoff that has the biggest statistically significant jump in enrollment using the Calonico et al. (2017) routine to generate estimates at each potential cutoff.[15] Visual evidence from these thresholds in Figure 1 shows a discontinuous jump in enrollment in fourth grade AWC by about 35 percentage points. As a check on this procedure, I also employ a split-sample strategy in which I estimate the cutoff in a randomly selected half of the sample and then generate estimates of treatment effects in the second half of the sample, repeating 10,000 times. Results from this exercise are generally consistent with my main findings, as are those from a test of placebo cutoffs, and are discussed in Section V and online Appendix A.

B. *The Fuzzy Regression Discontinuity*

A raw comparison of students who enroll in AWC with other BPS students would be misleading. AWC students are much more high achieving than the typical BPS student, and any difference in outcomes between the two groups could be due to underlying ability, rather than a program effect. Regression-based estimates of the AWC program that adjust for observable student characteristics like baseline test scores cannot fully address this problem; if there are unobserved differences between AWC students and other BPS students such as motivation or family interest in education, AWC effects would be confounded with omitted variable bias. To estimate the causal effect of AWC on students' outcomes unconfounded by omitted variable bias, I compare students just above and just below the eligibility thresholds to form regression discontinuity estimates of AWC's effect (Hahn, Todd, and Van der Klaauw 2001; Lee and Lemieux 2010). The only difference between students on either side of the threshold is the offer of AWC. The assumption here is that performance on a standardized test is a random draw from a student's underlying ability

---

private schools but choose to take the test to see if they qualify for AWC. These students are included in the calculation of distance to the threshold.

[15] The empirically derived thresholds are quite similar to the BPS thresholds in the years it is possible to compare the two, but not exactly the same, likely due to minor differences in data used to calculate the combined scores. The robustness checks include results using the derived cutoff for all cohorts, as well as the official cutoff only (where possible) and generally find similar results.

distribution, since students cannot precisely control their score on a test. Within a small window of points on an exam, students are in as-good-as random order, and the comparison between those above and below the threshold is analogous to the one in a randomized controlled trial.

The key assumption of regression discontinuity designs is that it is impossible to manipulate scores in order to qualify for the program (McCrary 2008). This assumption likely holds in the case of AWC. Since the threshold changes yearly, the exam is scored centrally, and students and teachers do not know the algorithm that translates questions answered correctly into exam scores, it is unlikely that students are able to manipulate their scores to qualify for the AWC.[16] This proves to be the case empirically. As shown in online Appendix Figure B.1, the frequency of test scores moves smoothly through the threshold, with no jump in frequency of a particular test score around the cutoff (panel A). Additional evidence of the smoothness of the distribution comes from a test suggested by Cattaneo, Jansson, and Ma (2017, 2018), shown in panel B. The *p*-value from that test is 0.801, suggesting no difference in density across the cutoff.

In a further check on the soundness of the regression discontinuity, I show that student background characteristics are smooth functions across the threshold in online Appendix Figure B.2 and confirmed with regressions in online Appendix Table B.2. Additionally, I use these covariates to generate predicted outcomes based on students beneath the threshold. Applying those predicted probabilities to all students is an approximation of what we would expect in the absence of the program. Graphs of these predicted outcomes are in online Appendix Figures B.3 through B.5 and show no discontinuities at the threshold, another piece of evidence that student characteristics are not what is driving differences across the threshold. An additional potential concern is that students differentially appear in the data based on their eligibility for AWC, perhaps with those above the threshold more likely to stay in the district and those just below to choose options like private schools. There is no differential attrition, as shown in online Appendix Table B.3.[17]

The AWC threshold is determined by a cut score for the combined math and reading scores, as described in Section IIIA. A measure of distance to the threshold, *Gap*, is the difference between the threshold and the combined score. Panel A of Figure 1 shows enrollment in fourth grade AWC by distance from the eligibility threshold. Adherence to the threshold rule is not perfect. A handful of students just below the threshold enroll in fourth grade AWC through the appeals process.[18] And a good proportion of students who qualify for the program do not take the offer, likely because it would involve switching schools or because they do not return their

---

[16] This is in contrast to the many gifted programs that admit students based on an IQ score threshold (McClain and Pfeiffer 2012), like the one studied in Card and Giuliano (2014, 2016). Since IQ scores have a subjective element, test administrators might shift students scores just above the threshold in order to give them access to gifted programming, either consciously or unconsciously.

[17] Given that some outcomes are missing for 15 to 30 percent of the sample, I conduct an exercise where I substitute predicted outcomes for missing data, which does not affect the interpretation of the results. Attrition is discussed in more detail in online Appendix A.

[18] Panel B of Figure 1 shows years of AWC enrollment by distance to the threshold. Since there is a second entry point to AWC in sixth grade, there appears to be more noncompliance since students below the threshold can accumulate years of AWC enrollment from the second entry point.

school choice forms. Thus to estimate the causal effect of AWC participation, I use a fuzzy regression discontinuity framework that accounts for imperfect compliance in a two-stage least squares (2SLS) setup. This is analogous to 2SLS estimates of causal effects in a randomized controlled trial with imperfect compliance. Estimates from this strategy will be local average treatment effects (LATEs) in two senses. First, results will be a weighted average treatment effect with weights proportional to the likelihood that a student will be in the "neighborhood" near the threshold (Lee and Lemieux 2010). Second, results will be local to compliers: those who attend AWC if their score passes the threshold and do not attend AWC if their score is below the threshold (Imbens and Angrist 1994; Angrist, Imbens, and Rubin 1996).

I model outcomes as a function of enrollment in the fourth grade AWC program. For a student i in the third grade in school s in school year t, I estimate a system of local linear regressions of the following form:

$$(1) \quad AWC_{ist+1} = \alpha_0 + \alpha_1 Above_{ist} + \alpha_2 Gap_{ist} + \alpha_3 Gap_{ist} \times Above_{ist} + \varepsilon_{ist},$$

$$(2) \quad Y_{ist+k} = \beta_0 + \beta_1 A\hat{W}C_{ist+1} + \beta_2 Gap_{ist} + \beta_3 Gap_{ist} \times Above_{ist} + \eta_{ist},$$

where Gap_{ist} measures distance to the AWC eligibility threshold on the third grade, $Above_{ist}$ is an indicator variable for being above the threshold in a given year, AWC_{ist+1} is indicator for enrollment in fourth grade AWC, and Y_{ist+k} is an outcome interest in some year, $t + k$, subsequent to third grade. The causal impact of AWC is represented by β_1 from the second stage regression, with program enrollment instrumented by program eligibility, $Above_{ist}$.

My preferred model estimates local linear regression with a triangular kernel of bandwidth 0.65 on either side of the program cutoff. The triangular kernel weights points near the threshold more heavily than those distant from the threshold. To settle on this bandwidth, I estimated optimal bandwidths for eight key outcomes[19] using the mean square error optimal bandwidths generated by the Calonico et al. (2017) and Calonico, Cattaneo, and Farrell (2018) procedure (henceforth referred to as the CCT bandwidths) and averaged the bandwidth across these eight outcomes to have a consistent sample across outcomes.[20] I later test the robustness of my findings to several additional bandwidths, including the CCT and Imbens and Kalyanaraman (2012) bandwidths computed for each outcome, and alternative specifications. Standard errors are clustered by third grade school by year.

The 2SLS estimates are the causal impacts of the program for compliers. I also report the control complier mean (CCM) as a measure of the mean of the outcome for students not eligible for the program. The CCM is the average outcome value for students underneath the threshold who are compliers, that is, those who accept the offer of AWC if they score high enough and do not attend AWC if they are below the cutoff, the population for whom the 2SLS procedure estimates an effect. The CCM

---

[19] The key outcomes are elementary school MCAS index; middle school MCAS index; an indicator for taking any AP; SAT score; indicators for high school graduation, college enrollment (any), and four-year college enrollment; and college quality. These are the same outcomes displayed in figures.

[20] This is an update to the routine described in Calonico, Cattaneo, and Titiunik (2014).

is not directly observable because those beneath the threshold who do not enroll in AWC are a mix of compliers and students who would never enroll in AWC even if eligible. Adapting the measurement of the CCM in the context of a randomized experiment in Katz, Kling, and Liebman (2001) to the fuzzy regression discontinuity setup using the methods discussed in Abadie (2002, 2003) yields the following equation for the CCM:

$$(3) \qquad Y_{ist+k} \times \left(1 - AWC_{ist+1}\right)$$

$$= \gamma_0 + \gamma_1\left(1 - A\hat{W}C_{ist+1}\right) + \gamma_2\,Gap_{ist} + \gamma_3\,Gap_{ist} \times Above_{ist} + \xi_{ist},$$

where $1 - AWC_{ist+1}$ is instrumented by AWC eligibility as in equation (2) and γ_1 is the estimate of the CCM. The CCM serves as my preferred measure of outcomes for the group beneath the threshold because a simple mean of students just below the threshold will commingle outcomes of compliers and non-compliers.

IV. Results

A. *First Stage and Effects on Enrollment and Attendance*

First-stage estimates of AWC enrollment are in Table 2. The first column shows a jump in enrollment of 36 percentage points at the threshold in fourth grade.[21] The first-stage *F*-statistic using enrollment in fourth grade AWC as the endogenous variable is 225. Just 3 percent of students just below the threshold enroll in fourth grade AWC, likely due to the appeals process. The next two columns show years of enrollment in fifth grade and sixth grade and above. I use initial enrollment in AWC in fourth grade as my endogenous predictor but report the jump in years of AWC attendance at additional grade levels since they are of interest themselves. By sixth grade, on average, students who are above the threshold enroll in AWC for an additional 0.79 years. Students who enroll in AWC tend stay in the program for about an additional 2.2 years (0.79/0.36) compared to those just below the threshold. In sixth grade and above, just below the threshold, the average number of years of AWC enrollment is 0.43. This increase in noncompliance is due to the second opportunity to enroll in AWC with a fifth grade test.

Like many urban school districts, BPS has faced declining enrollment since the 1970s, and since the introduction of charter schools in the late 1990s it must also now compete with the charter sector in Boston. AWC is one program that might draw families to the district or induce them to stay. Unlike other estimates of the effect of dedicated programs for high-achieving students on district enrollment (Figlio and Page 2002; Davis et al. 2013; Bui, Craig, and Imberman 2014), AWC qualification has few effects on the enrollment choices of students either during the grades that AWC serves or in subsequent grades, as shown in online Appendix Table B.4. Since enrollment in AWC by definition involves enrollment in BPS, I use

---

[21] Take-up of the AWC offer differs if a sending school houses an AWC program. The first stage in schools with a program is 45 percentage points; in other schools it is 34 percentage points.

TABLE 2—FIRST-STAGE ESTIMATES OF AWC ENROLLMENT

	4th grade (1)	5th grade (2)	6th grade and above (3)
Years AWC	0.356	0.662	0.788
	(0.023)	(0.046)	(0.065)
$\bar{Y}$	0.031	0.129	0.430
Observations	6,475	6,475	6,475

Notes: Each coefficient labeled "years AWC" is the estimate from an indicator for scoring above the AWC qualification threshold on years of AWC attendance for each grade level. All regressions include third grade year fixed effects. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.65. The sample is restricted to third graders enrolled in BPS in the fall of 2001 to 2005. Listed below each coefficient is the CCM. Robust standard errors clustered by baseline school by year are in parentheses.

Source: Author's calculations from BPS and DESE data

reduced form estimates based on eligibility to examine enrollment effects. There are no significant effects on enrollment in BPS, local charter schools, or other Massachusetts Public schools, and the magnitudes of the effects are quite small.[22] These are effects for students on the margin, however, and AWC may still have general equilibrium effects on enrollment.

AWC also does not significantly influence enrollment at Boston exam schools, which are three magnet schools for high achievers that also admit students based on test scores. This may be because a large majority of students are applying to an exam school anyway, as shown in online Appendix Table B.5.[23] AWC enrollment does appear to increase exam school offers in seventh grade by about 4 percentage points, but this difference is not statistically significant. These results mean that AWC does not achieve the goal of keeping families in the district or increasing the number of seats at exam schools that go to BPS students, at least for students on the margin of AWC eligibility.
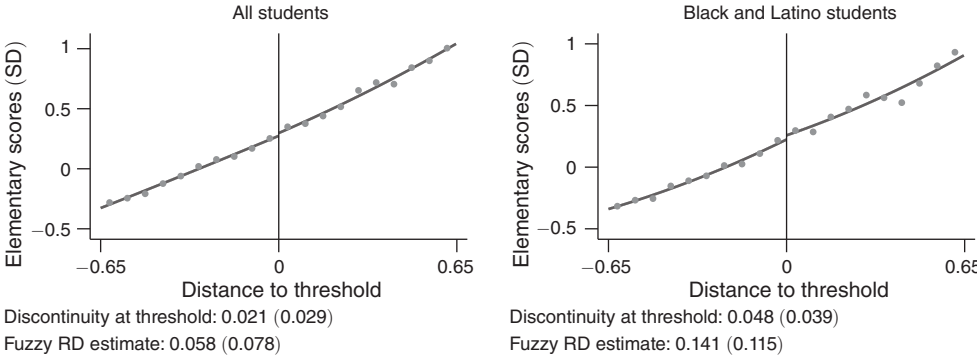
B. *MCAS Scores*

The first opportunity to examine the effect of the AWC program on academic outcomes is through MCAS scores, Massachusetts' standardized state tests, including exams offered in the grades in which AWC operates. Graphical evidence in

---

[22] While AWC does not increase the likelihood of enrollment in particular schools, it does increase attendance at school for black and Latino students, as shown in online Appendix Table B.11.

[23] The interaction between AWC enrollment and exam school application may have some explanatory power for the generally null results found in Abdulkadiroğlu, Angrist, and Pathak (2014). Seventy-two percent of students who enroll in AWC for at least 1 year apply to an exam school in seventh grade, with about 80 percent of those who applied receiving an offer. (These are descriptive findings, not from the RD sample.) About 36 percent of seventh grade exam school applicants have attended at least 1 year of AWC, and about 58 percent of grade exam school offers go to those who have enrolled in AWC. One of the reasons that exam schools appear to have little effect on student outcomes may be that a good number of exam school applicants have been treated by AWC. A difference may be that AWC occurs at a crucial point in time to give students a "foot in the door" and that later intervention cannot provide the same opportunity. On average, the RD sample students are between the cutoff scores for Boston Latin School, the most selective exam school, and Boston Latin Academy, the second most selective exam school.

Panel A. Elementary school scores

Discontinuity at threshold: 0.021 (0.029)
Fuzzy RD estimate: 0.058 (0.078)

Discontinuity at threshold: 0.048 (0.039)
Fuzzy RD estimate: 0.141 (0.115)

Panel B. Middle school scores

Discontinuity at threshold: 0.015 (0.028)
Fuzzy RD estimate: 0.041 (0.079)

Discontinuity at threshold: 0.055 (0.041)
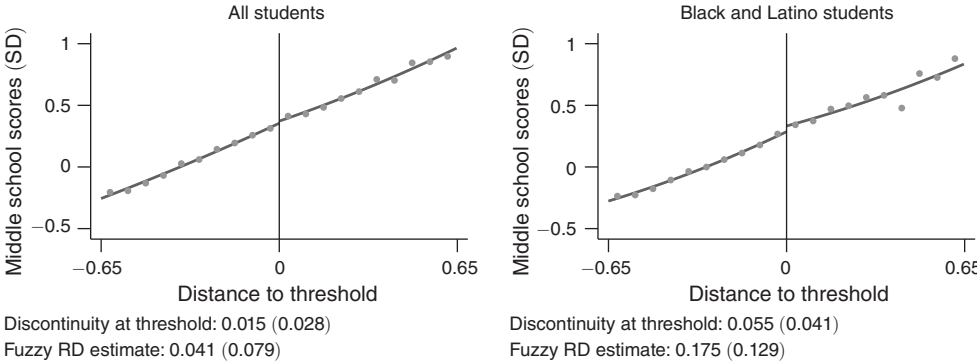Fuzzy RD estimate: 0.175 (0.129)

FIGURE 2. MCAS OUTCOMES BY DISTANCE TO ELIGIBILITY THRESHOLD

Notes: This figure shows average MCAS outcomes for bins of width 0.065 on either side of the threshold for all students (left side) and black and Latino students (right side) within the bandwidth of 0.65 around the eligibility threshold. A quadratic fit is imposed on either side of the threshold. Elementary school scores are the MCAS index for fourth and fifth grade students; middle school scores are the MCAS index for sixth, seventh, and eighth grade students.

Source: Author's calculations from BPS and DESE data

Figure 2 show no or very small jumps in MCAS scores at the threshold in both elementary and middle school. Fuzzy regression estimates corresponding to the figure are in Table 3. Stacking elementary school (fourth and fifth grades) and middle school (sixth through eight grades) outcomes increases precision, and I double cluster the standard errors from relevant regressions by student and third grade school by year. I combine test score outcomes (math and English/language arts, and science and writing in the grade levels where they are assessed) into one MCAS index, which is the standardized average of all subject *z*-scores in a grade, to reduce the possibility that significant results are chance findings due to multiple hypothesis testing. Results with scores by each subject are in online Appendix Table B.6, and results for additional subgroups are in online Appendix Table C.1.

AWC appears to generally increase standardized test scores in elementary and middle schools, as seen in Table 3, but the effects are imprecise. Enrollment in fourth grade AWC increases elementary school scores by 0.06σ and middle school

TABLE 3—FUZZY REGRESSION DISCONTINUITY ESTIMATES OF EFFECTS ON MCAS INDICES

	Elementary school (1)	Middle school (2)	10th grade (3)
Panel A. All students			
2SLS	0.058	0.041	0.149
	(0.078)	(0.079)	(0.113)
CCM	0.228	0.363	0.226
Observations (students)	6,161	5,741	5,209
Panel B. Black and Latino students			
2SLS	0.141	0.175	0.392
	(0.115)	(0.129)	(0.190)
CCM	0.151	0.233	−0.013
Observations (students)	3,837	3,523	3,208
Panel C. White and Asian students			
2SLS	−0.007	−0.066	−0.072
	(0.110)	(0.107)	(0.135)
CCM	0.310	0.490	0.489
Observations (students)	2,324	2,218	2,001

Notes: Each coefficient labeled 2SLS is the fuzzy regression discontinuity estimate of fourth grade AWC attendance on the outcome listed in the column heading. An indicator for scoring above the AWC qualification threshold is the instrument for AWC attendance. The specification uses local linear regression with a triangular kernel of bandwidth 0.65. Listed below each coefficient is the CCM. All regressions include third grade year fixed effects. The sample is restricted to third graders enrolled in BPS in the fall of 2001 to 2005. The MCAS index is the mean of all available MCAS subject-test z-scores, standardized to be mean zero, standard deviation one. Elementary school regressions stack fourth and fifth grade outcomes, include grade fixed effects, and double cluster standard errors by third grade school by year and student. Middle school regressions stack sixth, seventh, and eighth grade outcomes, include grade fixed effects, and double cluster standard errors by third grade school by year and student.

Source: Author's calculations from BPS and DESE data

scores by 0.04σ. Recall that the test scores are standardized to the state to mean zero and standard deviation one so that control compliers generally have positive scores. Test score impacts are larger for black and Latino students but still have large standard errors. One reason why it may be difficult to detect impacts on test scores is that there is a ceiling effect with high-achieving students who make up the RD sample "topping out" on the MCAS. This is not the case. Very few students score at the very top of the exam, and there is no differential effect on top scoring by AWC participation (column 6 of online Appendix Table B.6). Test score impacts are larger in tenth grade and are statistically significant for black and Latino students. Across various bandwidths and alternative specifications, as shown in online Appendix Tables A.1 and A.2, impacts on test scores are generally positive, larger for black and Latino students, but only occasionally statistically significant.

If what matters for academic achievement is relative position in the academic distribution, as posited by Marsh (1987) (the "big-fish-little-pond effect") and shown to have a role by Elsner and Isphording (2017) and Murphy and Weinhardt (2018), an investigation of whether or not AWC influences class rank is also relevant. Thus, I also show the effects of AWC on MCAS-based class rank within a school in online

Appendix Table B.7. Attendance at AWC generally decreases class rank, due to the larger concentration of high-achieving students in schools with AWC. While class rank decreases are typically not statistically significant, if anything they would imply a decrease in achievement, as in Murphy and Weinhardt (2018), the opposite of what is seen here.

Embedded in the tenth grade MCAS exam are two additional milestones, also shown in online Appendix Table B.7. For the class of 2003 forward, students in Massachusetts have had to meet a competency threshold on the high school MCAS exam in order to graduate. AWC increases the likelihood that students meet this determination by a significant 11 percentage points (a marginally significant 16 percentage points for black and Latino students). At a higher threshold on the tenth grade MCAS, students become eligible for the Adams Scholarship, which pays for tuition at public institutions in Massachusetts.[24] AWC compliers are 13 percentage points more likely to qualify for the Adams Scholarship, though this finding is not statistically significant ($p = 0.11$).[25] Passing these thresholds does not guarantee high school graduation or college enrollment; however, both may serve as positive feedback to students that they are on track academically.

C. *Other Academic Outcomes*

Standardized test scores only tell a partial story in terms of academic potential. In Table 4, I present estimates for key middle and high school outcomes that are related to success in higher education and in general: Algebra 1 enrollment by eighth grade, AP, SAT, and high school graduation. One of the goals of the AWC program is to accelerate math. Column 1 of Table 4 shows a marginally significant 25 percentage point increase in enrollment in Algebra 1 by eighth grade, suggesting that AWC is successful at this goal.[26]

A graphical presentation of the discontinuity is in Figure 3 for AP test-taking, SAT scores, and high school graduation. AP courses are an important part of higher education preparation. They offer an opportunity for rigorous course experiences as well as potential college credit. Jackson (2014) shows that incentive payments for AP exam scores increased AP scores, college persistence, and wages, demonstrating the possibility of AP as an important mediator for adult outcomes. AWC participants are more likely than their counterparts to take an AP exam, with a 12 percentage point increase in exam participation ($p = 0.16$). Below the threshold, 54 percent of control complier students take an AP exam. Thus attendance at AWC increases participation in the AP program by 22 percent. Unlike many of the other results

---

[24] Cohodes and Goodman (2014) finds that this scholarship induces students to lower quality institutions than they would have otherwise attended, thus decreasing graduation rates. However, this finding is driven by relatively higher income students, unlike the BPS students examined here.

[25] There is some increase in enrollment in Adams institutions in this context. AWC increases enrollment in Adams-eligible institutions by 11 percentage points, though most of that enrollment increase is in the 2-year sector, and the change is not statistically significant.

[26] Data on course enrollment is only available for more recent cohorts of data, and thus the Algebra 1 estimate is generated in the third grade cohorts from 2005 to 2008, not the same sample used elsewhere. Estimates for higher level math courses (not shown) do not show a consistent pattern of increased enrollment, which may be due to different math course sequences across high schools.

TABLE 4—FUZZY REGRESSION DISCONTINUITY ESTIMATES OF EFFECTS ON ACADEMIC OUTCOMES

	Algebra 1 by 8th grade (1)	Took any AP (2)	Number of APs taken (3)	Took SAT (4)	SAT score (5)	On-time high school graduation (6)	Late high school graduation (7)
Panel A. All students							
2SLS	0.254	0.118	0.075	0.058	−43.662	0.081	0.034
	(0.131)	(0.085)	(0.354)	(0.059)	(38.057)	(0.059)	(0.044)
CCM	0.505	0.540	1.546	0.820	1,565.433	0.810	0.912
Observations	4,456	4,671	4,671	4,671	3,844	4,671	3,771
Panel B. Black and Latino students							
2SLS	0.240	0.094	0.141	0.127	9.330	0.207	0.099
	(0.186)	(0.127)	(0.455)	(0.094)	(49.072)	(0.095)	(0.071)
CCM	0.487	0.523	1.376	0.736	1,489.348	0.665	0.818
Observations	2,911	2,814	2,814	2,814	2,173	2,814	2,253
Panel C. White and Asian students							
2SLS	0.284	0.211	0.378	0.004	−57.714	−0.055	−0.026
	(0.196)	(0.105)	(0.535)	(0.072)	(54.571)	(0.058)	(0.050)
CCM	0.556	0.529	1.718	0.900	1,630.521	0.980	1.006
Observations	1,545	1,857	1,857	1,857	1,671	1,857	1,518

Notes: Each coefficient labeled 2SLS is the fuzzy regression discontinuity estimate of fourth grade AWC attendance on the outcome listed in the column heading. An indicator for scoring above the AWC qualification threshold is the instrument for AWC attendance. The specification uses local linear regression with a triangular kernel of bandwidth 0.65. Listed below each coefficient is the CCM. All regressions include third grade year fixed effects. The sample is restricted to third graders enrolled in BPS in the fall of 2001 to 2005. Robust standard errors clustered by third grade school by year are in parentheses. For Algebra 1 by eighth grade, the sample includes students who match to the student course data (2011–2014), which are the fall cohorts from 2005–2008. On-time high school graduation is an indicator for high school graduation 10 years after the third grade exam for AWC eligibility; late high school graduation is an indicator for 11 years after.
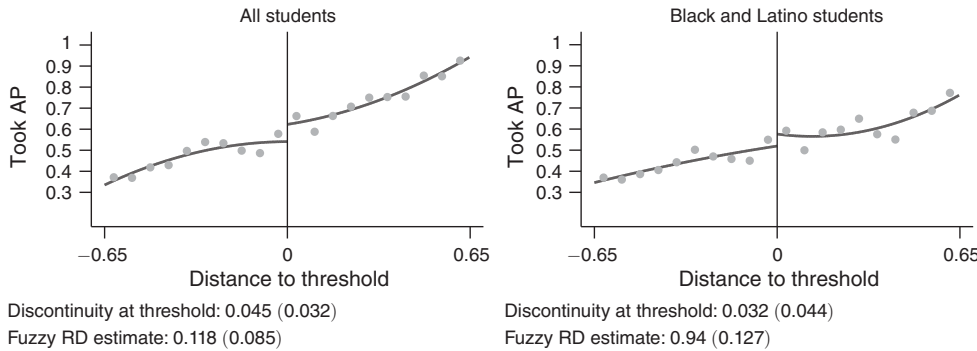
Source: Author's calculations from BPS and DESE data

in this paper, the AP gains are concentrated among white and Asian students. The increase in AP exam taking comes mostly from a significant 11 percentage point increase in students taking AP economics.[27] Robustness checks in online Appendix Tables A.1 and A.2 show that the AP gains are occasionally marginally statistically significant in some specifications for the full sample but generally imprecise.
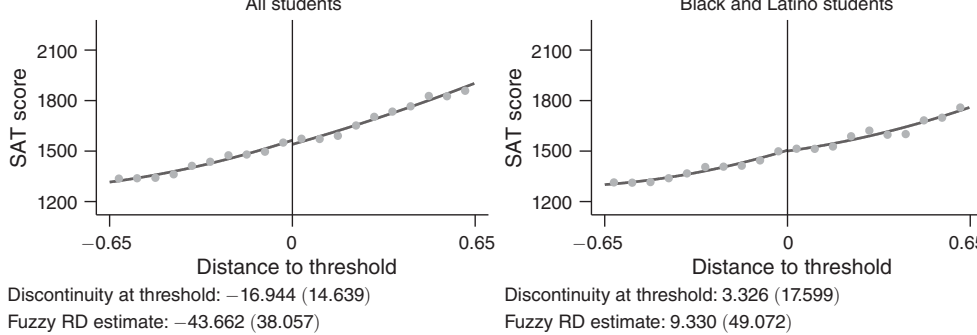
Taking the SAT is another key milestone for application to college, as many four-year colleges require the exam.[28] Policies that induce universal access to college entrance exams also increase college going, indicating that not taking the exam is a barrier (Hyman 2017, Hurwitz et al. 2015). As seen in Table 4, about 82 percent of control compliers take the SAT, and AWC increases the test-taking rate by about 6 percentage points, though this result is not significant. Gains in SAT taking are driven by black and Latino students, who are 13 percentage points more likely to take the test, off a lower control complier rate of 73 percent ($p = 0.18$).

---

[27] For subject-specific AP results, see online Appendix Table B.9.
[28] Colleges also accept the ACT, but most students in Massachusetts take the SAT.

Panel A. Took any AP exam

Discontinuity at threshold: 0.045 (0.032)
Fuzzy RD estimate: 0.118 (0.085)

Discontinuity at threshold: 0.032 (0.044)
Fuzzy RD estimate: 0.94 (0.127)

Panel B. SAT score (2400)

Discontinuity at threshold: −16.944 (14.639)
Fuzzy RD estimate: −43.662 (38.057)

Discontinuity at threshold: 3.326 (17.599)
Fuzzy RD estimate: 9.330 (49.072)

Panel C. Graduated high school on time

Discontinuity at threshold: 0.031 (0.022)
Fuzzy RD estimate: 0.081 (0.059)

Discontinuity at threshold: 0.071 (0.032)
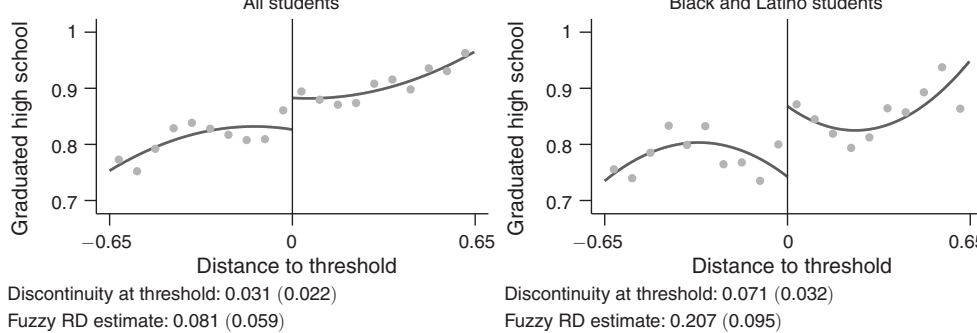Fuzzy RD estimate: 0.207 (0.095)

Figure 3. High School Outcomes by Distance to Eligibility Threshold

Notes: This figure shows average high school outcomes for bins of width 0.065 on either side of the threshold for all students (left side) and black and Latino students (right side) within the bandwidth of 0.65 around the eligibility threshold. A quadratic fit is imposed on either side of the threshold.

Source: Author's calculations from BPS and DESE data

There is no effect on SAT scores for those that take the exam.[29] Given that taking the SAT is a prerequisite for application to colleges that are not open enrollment, the

---

[29] See online Appendix Table B.10 for subject-specific results.

gains for black and Latino students are particularly notable as they open a path to four-year college AWC increases SAT-taking rates for black and Latino students to the level of their white and Asian counterparts. Online Appendix Table C.2 shows that the SAT-taking gains are even larger, and statistically significant, for students who had lower MCAS scores in third grade, increasing test-taking rates from 64 percent to 89 percent.[30]

AWC has a positive, nonsignificant effect on high school graduation overall of 8 percentage points, with on-time graduation year calculated based on third grade exam year.[31] The increase in on-time high school graduation for minority students is large, with a statistically significant gain in graduation rate of 21 percentage points. This implies that AWC brings up the graduation rate for black and Latino students to 88 percent, given that the graduation rate for compliers just beneath the threshold is low, at 67 percent. Robustness checks indicate that this finding is consistent across most specifications. About half of this increase in high school graduation comes from those who would not graduate in absence of the program, as there is an increase in the five-year high school graduation rate of 10 percentage points for black and Latino students ($p = 0.16$), as shown in the final column of Table 4. The high school graduation gains are concentrated among lower MCAS scoring students, as shown in online Appendix Table C.2.

On-time graduation is made possible, in part, by two factors that AWC may also influence. AWC increases the likelihood that students will appear in twelfth grade on time by 8 percentage points ($p = 0.18$) for black and Latino students (see online Appendix Table B.12). It also increases the likelihood that students meet the MCAS graduation requirement in their first attempt (see online Appendix Table B.7). I combine these two outcomes into a single indicator for achieving *both* outcomes in the final column of online Appendix Table B.12. AWC increases the likelihood of jointly meeting these milestones by 13 percentage points for all students and 19 percentage points for black and Latino students, indicating that AWC puts students on a trajectory that sets the stage for graduating high school on time, making it possible to enroll in college on time.

D. *College*

The AWC program begins almost a decade before college enrollment, but it has a long-lasting impact on students' college behavior. Students who participate in AWC are more likely to enroll in college the fall after expected high school graduation, as seen in column 1 of Table 5. Fifty-three percent of compliers just below the threshold enroll in college on time, with a marginally significant increase in enrollment of 15 percentage points for AWC attendees, a 28 percent increase. While most of the robustness checks show findings of similar magnitudes, they are not consistently statistically significant, as discussed more in Section V, Part A. Results

---

[30] I define "high MCAS students" as those who score 0.25σ or higher on their third grade ELA MCAS; lower MCAS students are those who score below that threshold. Since AWC eligibility is determined by an alternative exam, both high- and lower MCAS students qualify for the program. About one-third of students in the RD sample are lower MCAS students, who still have relatively high scores compared to BPS as a whole.

[31] Note that on-time graduation is typically calculated from ninth grade enrollment.

Table 5—Fuzzy Regression Discontinuity Estimates of Effects on College Enrollment

	On-time college enrollment			College quality in 2014 US\$ (4)	Late enrollment any (5)
	Any (1)	Four-year (2)	Two-year (3)		
Panel A. All students					
2SLS	0.149	0.063	0.087	1,788	0.169
	(0.080)	(0.081)	(0.045)	(3,866)	(0.084)
CCM	0.532	0.522	0.009	43,654	0.548
Observations	5,502	5,502	5,502	5,502	4,567
Panel B. Black and Latino students					
2SLS	0.262	0.182	0.081	8,224	0.246
	(0.116)	(0.120)	(0.071)	(5,456)	(0.118)
CCM	0.402	0.383	0.018	38,351	0.465
Observations	3,374	3,374	3,374	3,374	2,802
Panel C. White and Asian students					
2SLS	0.059	−0.031	0.090	−4,689	0.071
	(0.106)	(0.107)	(0.054)	(5,217)	(0.108)
CCM	0.686	0.689	0.000	51,771	0.679
Observations	2,128	2,128	2,128	2,128	1,765

Notes: Each coefficient labeled 2SLS is the fuzzy regression discontinuity estimate of fourth grade AWC attendance on the outcome listed in the column heading. An indicator for scoring above the AWC qualification threshold is the instrument for AWC attendance. The specification uses local linear regression with a triangular kernel of bandwidth 0.65. Listed below each coefficient is the CCM. All regressions include third grade year fixed effects. The sample is restricted to third graders enrolled in BPS in the fall of 2001 to 2005. Robust standard errors clustered by third grade school by year are in parentheses. On-time college entrance is calculated based on entry into college 10 years after the third grade exam for AWC eligibility, late enrollment based on entry 11 years after the exam. College quality earnings outcomes are measured by the estimated 2014 earnings of college attendees from the 1980–1982 birth cohorts from Chetty et al. (2017). Students are assigned the earnings outcomes of the college they attend by gender even if they are not on-time attendees. Students who do not attend college are assigned the outcomes for non-attendees of the same gender.

Source: Author's calculations from BPS, DESE, and NSC data

in column 5 showing "late enrollment," which includes an additional year after projected high school graduation, are very similar, indicating that AWC increases college participation rather than shifting its timing. With a similar CCM of 53 to 55 percent enrolling, this also indicates that the bulk of enrollment gains are coming from on-time high school graduates and that late high school graduation, even if this milestone is completed, does not necessarily translate into college enrollment at the same rate that on-time graduation does. The college enrollment effect comes from increased matriculation at both four- and two-year institutions, with each accounting for about half of the enrollment gains.

Black and Latino students are again driving these overall impacts, with even larger AWC effects for this sample. For these students, the on-time enrollment increase is 26 percentage points, a 65 percent increase in enrollment over the 40 percent enrollment rate for control compliers. This finding is consistent across most robustness checks. Over two-thirds of the increase for black and Latino students comes from attendance at four-year universities.

To consider the quality of the institutions attended, I link students' attended colleges with the mean estimated earnings of attendees of that college in 2014 dollars, as estimated in Chetty et al. (2017).[32] Students are assigned the earnings associated with their attended college by gender, even if they do not enroll on time (since Chetty et al. (2017) does not distinguish between on-time and later enrollment), and non-attendees are assigned the non-attendee estimated earnings from Chetty et al. (2017), also by gender.[33] I consider the estimated earning variables a measure of college quality, as they demonstrate the real-world differences in outcomes across colleges, and they should not be interpreted as implying that an attendee at a college is guaranteed that level of earnings.

For all students, attending AWC increases college quality by a little more than $1,750, though this difference is not significant. Gains in college quality are driven by black and Latino students, where AWC induces a college quality increase of about $8,200 ($p = 0.13$). This shifts average college quality on par with the colleges attended by white and Asian students.[34] While this exercise cannot predict the actual earnings gains from attending AWC, it is a useful exercise to show the life-changing potential of attending college, which is a major effect of participating in AWC. However, college quality estimates for black and Latino students only become marginally statistically significant with larger bandwidths (online Appendix Table A.2). A graphical presentation of these outcomes is in Figure 4.

The change in distribution of college quality due to AWC enrollment is also informative, and differences are statistically significant here even when they are not at the mean. I estimate the densities of college quality distributions for compliers, using an adaptation of the methods for estimating the CCM, so that I interpret distributional differences as causal effects free from selection bias. Specifically, complier densities are estimated over a grid of 100 points using a Gaussian kernel and Silverman (1986) rule-of-thumb bandwidth.[35] I also report Kolmogorov-Smirnov statistics, which are maximum differences in complier CDFs, and bootstrap p-values for this difference.

Figure 5 shows how AWC shifts students out of the left tail of the distribution, where non-attendees make up the bulk of the left tail for untreated compliers (dashed line). Panel A shows the distributions for all students (bootstrapped p-value = 0.11). The shift away from nonattendance is especially large for black and Latino students (panel B), and their enrollment in college can be seen in the increased density for treated compliers (solid line) from about $40,000 to about $90,000. The distributions in panel B for untreated and treated compliers are significantly different from each other (bootstrapped p-value = 0.02). Enrollment in elite institutions (college quality of $100,000 and up) remains very similar for black and Latino treated and untreated

---

[32] An alternative measure of quality is the Barron's rankings of selectivity. Enrollment increases are largest in the Barron's categories Chetty et al. (2017) categorizes as "selective" (Barron's rating 3 through 5). However, this finding closely parallels the increase in four-year enrollment so that Barron's rankings do not provide additional information.

[33] Note that Chetty et al. (2017) is unable to distinguish between campuses of the University of Massachusetts, though they do have separate earnings estimates for state and community colleges.

[34] Online Appendix Table C.4 shows estimates on college outcomes for additional subgroups.

[35] See Angrist et al. (2016); Abdulkadiroğlu, Pathak, and Walters (2018); and Walters (2018) for implementations of such densities in other contexts and details on estimation. I update the methods used in these papers to the regression discontinuity setup.

Panel A. Enrolled any college on time
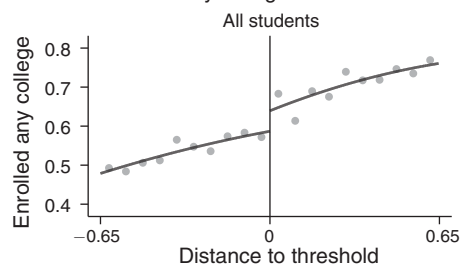
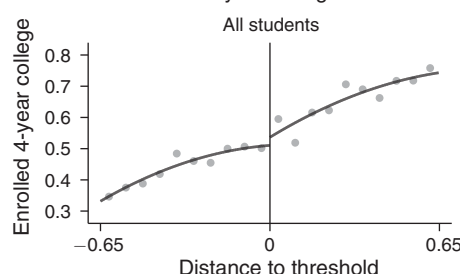Discontinuity at threshold: 0.055 (0.030)
Fuzzy RD estimate: 0.149 (0.080)
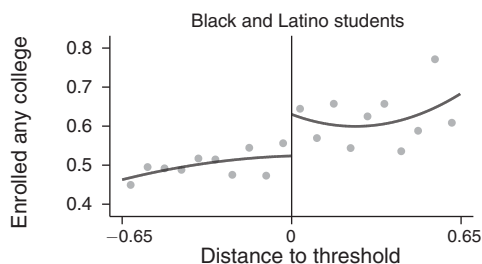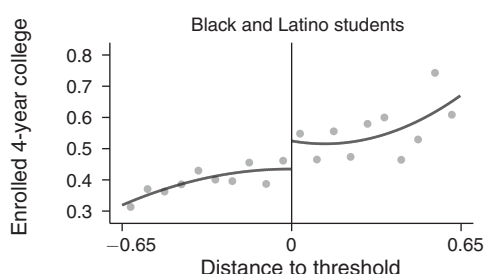
Discontinuity at threshold: 0.089 (0.039)
Fuzzy RD estimate: 0.262 (0.116)

Panel B. Enrolled four-year college on time
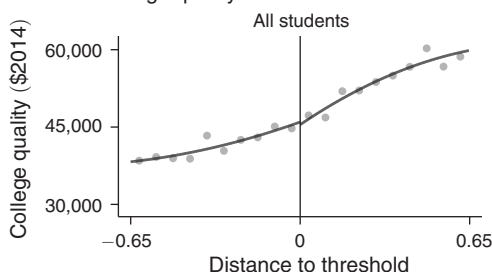

Discontinuity at threshold: 0.023 (0.03)
Fuzzy RD estimate: 0.063 (0.081)

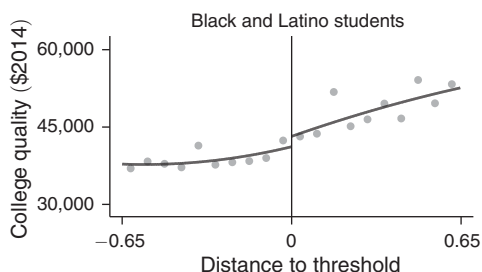Discontinuity at threshold: 0.062 (0.041)
Fuzzy RD estimate: 0.182 (0.120)

Panel C. College quality

Discontinuity at threshold: 663 (1,436)
Fuzzy RD estimate: 1,788 (3,866)

Discontinuity at threshold: 2,786 (1,861)
Fuzzy RD estimate: 8,224 (5,456)

Figure 4. College Enrollment Outcomes by Distance to Eligibility Threshold

Notes: This figure shows average college enrollment outcomes for bins of width 0.065 on either side of the threshold for all students (left side) and black and Latino students (right side) within the bandwidth of 0.65 around the eligibility threshold. A quadratic fit is imposed on either side of the threshold. College quality earnings outcomes are measured by the estimated 2014 earnings of college attendees from the 1980–1982 birth cohorts from Chetty et al. (2017). Students are assigned the earnings outcomes of the college they attend by gender even if they are not on-time attendees. Students who do not attend college are assigned the outcomes for non-attendees of the same gender.

Source: Author's calculations from BPS, DESE, and NSC data

compliers, showing that most of the increase in enrollment comes from non-elite four-year institutions. This includes University of Massachusetts campuses, state universities and colleges, and private nonprofit institutions like Suffolk University and Wentworth Institute of Technology. Thus, increases in college quality come mostly from shifting from nonattendance to nonselective or somewhat selective institutions.
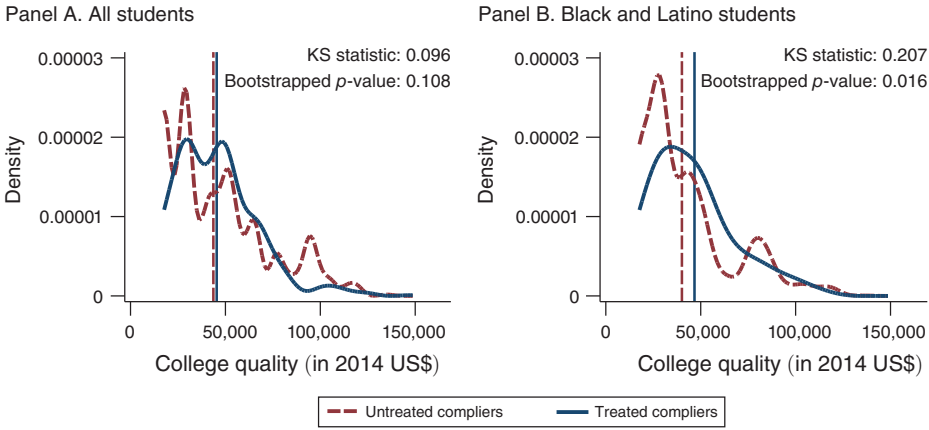
Panel A. All students

Panel B. Black and Latino students

FIGURE 5. COLLEGE QUALITY DISTRIBUTIONS FOR TREATED AND UNTREATED COMPLIERS

Notes: This figure shows college quality distributions for AWC compliers. All models use a Gaussian kernel and the Silverman (1986) rule-of-thumb bandwidth. Vertical dashed lines indicate CCM college quality, and solid lines indicate treated complier mean college quality. Kolmogorov-Smirnov statistics are maximum differences in complier CDFs, and *p*-values are bootstrapped. College quality earnings outcomes are measured by the estimated 2014 earnings of college attendees from the 1980–1982 birth cohorts from Chetty et al. (2017). Students are assigned the earnings outcomes of the college they attend by gender even if they are not on-time attendees. Students who do not attend college are assigned the outcomes for non-attendees of the same gender.

Source: Author's calculations from BPS, DESE, and NSC data

E. *Magnitude of the Effects*

The magnitudes of the AWC gains described here are large, in some cases quite large. One reason for this is that control complier attainment is low. Black and Latino control compliers have a 67 percent on-time high school graduation rate and 40 percent on-time college enrollment rate. The low counterfactual and relatively high-achieving student body leaves room for large potential increases. Another is that the estimates are LATEs, and marginal students and compliers may benefit more than the average student from the AWC intervention. Additionally, the long time period between the treatment and outcomes I focus on leave many potential opportunities to fall off track. Rather than boosting students above and beyond, AWC appears to maintain their status, preventing black and Latino students from a "leaky pipeline." These small gains may build upon each other, with AWC providing the crucial "foot in the door" that begins a chain of positively reinforcing events (Bailey et al. 2017).

Figure 6 summarizes this pipeline. It displays treated and untreated complier means for on-time enrollment in middle and high school and on-time high school graduation and college enrollment separately for white and Asian students and black and Latino students.[36] All of the groups of students maintain on-time progress through school until tenth grade, but the lines begin to separate in eleventh grade.

---

[36] Each point estimate for each group is generated by a separate regression using the main fuzzy regression discontinuity setup in the sample of students who appear in the data for all of the outcomes included in the graph,
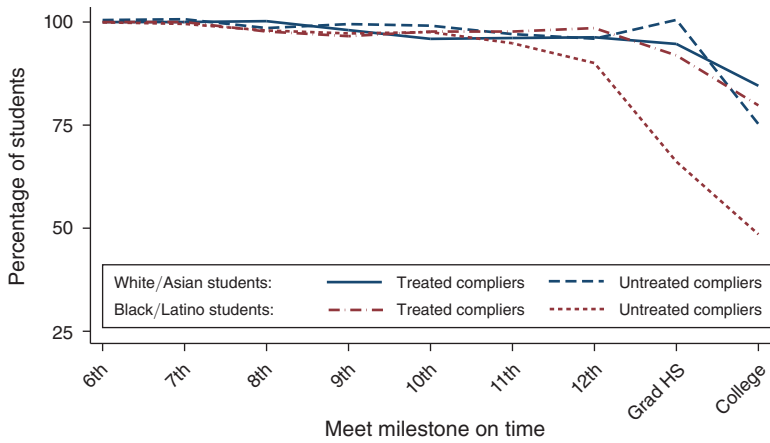
FIGURE 6. PROGRESS THROUGH SCHOOL FOR TREATED AND UNTREATED COMPLIERS

Notes: This figure plots the CCMs and treated complier means for white and Asian students and black and Latino students for on-time progress measured through school. Means for each outcome and group are estimated using the standard fuzzy regression discontinuity specification, as described in equation (3). Since there are no differences in fourth and fifth grade outcomes, they are omitted. The sample is limited to students who are present in the data for all of the outcomes.

Source: Author's calculations from BPS, DESE, and NSC data

Untreated black and Latino compliers begin to fall away from the other groups, achieving on-time progress at lower rates. The gap continues to widen in twelfth grade and on-time high school graduation. Notably, black and Latino AWC compliers continue to meet milestones and do so at the same rate as their white and Asian peers (where there are few differences by treated status). This figure makes clear that the AWC treatment helps students to maintain success in progress through school, and having just missed the cutoff for the treatment puts students at risk for falling "off track." While the enrollment and high school graduation effects are large, they do not occur in isolation. Instead, they rely on sequential increases in meeting milestones along the way.

The role of on-time progress through school in supporting high school graduation has been found in other contexts as well. Mariano, Martorell, and Tsai (2018) finds that test score cutoff-induced grade retention in seventh and eighth grade in New York City increases dropout rates by 10 percentage points and decreases on-time graduation rates by 30 percentage points. Investigating a similar policy of test-induced retention, Eren, Depew, and Barnes (2017) finds that grade retention increases high school dropout by 5 percentage points. Using competency requirements for high school graduation embedded in the tenth grade MCAS exam, Papay, Murnane, and Willett (2010) finds that low-income urban students that just miss the cutoff for passing the math exam are 8 percentage points less likely to graduate high school.

---

using equation (3) for untreated compliers and adapting that equation for treated compliers by using an indicator for AWC enrollment rather than $1 - AWC$.

Other interventions that help students stay "on track" can also have large effect sizes. In a scenario similar to AWC, Dougherty et al. (2017) examines the effect of using a test to determine math placement (rather than subjective measures). This policy induces students who would not otherwise be placed in an accelerated math course to be placed in a college preparatory sequence. Students increase their on-time progress and completion of math courses, and their intention to attend a four-year college goes up by 25 to 28 percentage points. This treatment effect is on par with the AWC effect estimated here, however intentions may not translate to actual enrollment. A small randomized controlled trial found that self-affirmation writing exercises in seventh grade increased enrollment in college by 22 percentage points for black students—quite a large effect considering the low-intensity intervention (Goyer et al. 2017). The offer of AWC and the program itself may similarly serve as self-affirmation.

Other interventions that share some similarities with AWC can have large effects, though they are less closely aligned with AWC as the studies described above. Dee and Penner (2017) analyzes the effect of being assigned to an ethnic studies class for students below a grade point average (GPA) threshold and finds large increases in attendance, GPA, and course credits. While this program focused on low-scoring students, the treatment of culturally relevant pedagogy may also induce positive identity formation, a potential mechanism in the AWC case as well. Dougherty (2018) finds that career and technical education (CTE) high schools for higher income and more white populations increases high school graduation rates by 17 to 35 percentage points for compliers. Intensive tutoring for high school boys in Chicago resulted in much higher math scores and grades and a 24 percentage point jump in an indicator for being "on track" (Cook et al. 2014). Cook et al. (2014) estimates that the "on-track" indicator would translate into a 14 percentage point increase in high school graduation if study participants follow similar patterns to other Chicago students. Comprehensive support for low-income youth in housing programs increased high school graduation by 15 percentage points and college enrollment by 19 percentage points (Oreopoulos, Brown, and Lavecchia 2017). In both these cases interventions provided access to more intensive educational experiences, similar to AWC, but participating students were generally much lower achieving than AWC students. Barrow, Schanzenbach, and Claessens (2015) shows that small schools in Chicago increase high school graduation by 20 percentage points. Attendance at Harlem Children's Zone Promise Academy increases on-time high school graduation by 22 percentage points and on-time college enrollment by 28 percentage points (Dobbie and Fryer 2015). Boston charters increase four-year college enrollment by 13 to 18 percentage points, though they do not increase (and may delay) on-time high school graduation. As a whole, the effects of AWC are large but not unheard of.

V. Threats to Validity

A. *Robustness Checks*

In online Appendix Table A.1, I present results for eight key outcomes for a variety of specifications and bandwidths for all students. Online Appendix Table A.2 has

analogous robustness checks for the black and Latino student population, as AWC effects occur most consistently for these students. Panel A replicates my default specification for reference purposes in the first line and then is followed by a number of alternative specifications. The alternatives are including baseline covariates, including school by year fixed effects, using only the empirically derived cutoffs, using only the official cutoffs (which limits the sample to cohorts from 2003 and beyond), using a quadratic functional form on the full sample, and using predicted outcomes for missing data to address concerns about attrition (discussed in detail in online Appendix A). Panel B reports estimates for four set bandwidths at 0.25, 0.5, 0.75, and 1. It also includes estimates from the standard specification estimated on the CCT and IK bandwidths, which are generated separately for each outcome and sample. The details on various bandwidths are presented in online Appendix A and generally follow a pattern of larger but less precise impacts for smaller bandwidths and smaller but more precise impacts for larger bandwidths.

As discussed above, the results are generally robust to a number of specification checks, especially for black and Latino students. Results from the original specification with covariates are very similar. Students are not restricted by policy to a particular AWC program, however the location of their school may restrict their choice set. Including school by year fixed effects, as in the third row of panel A, accounts for this by restricting comparisons to within third grade sending school and year. Estimates for black and Latino students on high school graduation and college enrollment are slightly smaller and now marginally significant. Findings for all students are affected more. The inclusion of fixed effects reduces the enrollment and college quality estimates considerably, and they are no longer statistically significant. This is due to the fact that estimates for white and Asian students with fixed effects are a small, nonsignificant negative. In other words, comparing white and Asian students within the same sending schools yields a small negative, but comparing those students across and within yields a small positive. Since it is black and Latino students who are driving the effects for all students throughout the paper, and the estimates for these students remain statistically significant with fixed effects, the reduction in the estimates for all students is less concerning.

As noted in Section III, Part B, I have official cutoff scores from BPS for the 2003 third grade cohort forward and thus must use empirically derived cutoffs for the 2001 and 2002 cohorts. Results using the empirically derived cutoffs for all cohorts are shown in the rows labeled "derived cutoffs." For all students, the findings are generally similar, and the college enrollment coefficient becomes nonsignificant. For black and Latino students, using only derived cutoffs yields smaller college effects (19 percentage points), and the college enrollment estimate becomes marginally significant. However, the high school graduation effect is actually larger and more precise with derived cutoffs (23 percentage points). Limiting the sample to cohorts with official cutoffs from 2003 and later in the next row results in high school graduation and college enrollment estimates that are generally of similar magnitude for black and Latino students but less precise. The decrease in precision is likely due to the decrease in sample size that comes from removing two cohorts from the analysis. However, the estimate for college quality is much smaller. One limitation of this study is the lack of official cutoffs for the first two cohorts analyzed. These

additional specifications provide some evidence around alternative thresholds, and I examine the cutoffs further in Section V, Part B below.

I also fit quadratic polynomials of the running variable on either side of the threshold, using the whole sample and no weights, with results displayed in the second to last row of panel A, labeled "quadratic."[37] The parametric approach yields results in the same direction, though generally smaller. However, the college enrollment estimate remains marginally statistically significant for black and Latino students. When I employ a quadratic specification but limit the bandwidth, results are very similar to the main estimates (not shown).

Overall, the bandwidth and specification details are generally consistent with the main findings, with some exceptions. Notably, the gains in college enrollment for black and Latino students are of similar magnitudes in all of the robustness checks and statistically significant in most.

B. *Placebo Cutoffs*

Since cutoffs are key to identification in this scenario, I provide an additional piece of evidence based on the cutoffs.[38] Online Appendix Figures A.1 to A.3 show the distribution of regression discontinuity estimates at placebo thresholds. I include all possible cutoffs incremented by 0.01 where there are at least 50 students on either side of the threshold. Compliance is not interpretable away from the threshold, thus this exercise uses regression discontinuity estimates of an offer of AWC, not the fuzzy RD.

The figures plot the distribution of the reduced form estimates, with the standard threshold estimate indicated by a dashed line. An estimate at the 97.5 percentile or above is a statistically significant result with a nonparametric 95 percent confidence interval; similarly, an estimate at the ninety-fifth percentile or above is statistically significant at a nonparametric 90 percent confidence interval. Given the imprecise estimates for MCAS scores, it is not surprising that results from this placebo test are in the middle of the distribution, as seen in online Appendix Figure A.1. For high school outcomes in online Appendix Figure A.2, the AP test result for all students is at the 92.6th percentile. The high school graduation estimate was particularly notable for black and Latino students. Here, it corresponds to the 92.5th percentile, which would only be significant with a one-tailed test and a 90 percent nonparametric confidence interval. College enrollment estimates, in online Appendix Figure A.3, are at the 94.5th percentile for all students (just missing the threshold for significance with a 90 percent confidence interval) and 97.9th percentile for black and Latino students (meeting the threshold for statistical significance with a 95 percent confidence interval). The estimate for college quality for black and Latino students is at the 94.3th percentile. The findings from this exercise are generally consistent with the main conclusions described above.

---

[37] Following Gelman and Imbens (2019), I do not estimate parametric models with higher order polynomials.

[38] Online Appendix A presents a second piece of evidence on the cutoffs. To address the concern that estimating empirical thresholds on the same sample that I estimate outcomes on introduces bias, I conduct a split sample test where half of the data is used to estimate the thresholds and the other half is used to estimate program effects.

VI. Mechanisms

A. *Differences between AWC and Non-AWC Classrooms*

In the estimates above, I have not specified a specific channel through which the AWC program boosts longer-term outcomes. It could be some specific aspect of the program, the positive identity that the program helps develop, or that AWC sets students on an accelerated track that make on-time college enrollments possible. This section will discuss potential mechanisms, first documenting that there is a difference in classroom experiences between AWC and non-AWC classrooms. In Table 6, I show that AWC classrooms are different than the alternate classrooms attended by those just below the threshold. These results for fourth through sixth grade classroom characteristics are limited to more recent years of data since that is when student-teacher-course links are available in the state data.[39] Columns 1 through 5 show that AWC alters classroom composition, as measured by demographic and other third grade characteristics. As first shown observationally in Table 1, the causal effect of AWC on demographic composition is fewer black and Latino students in the classroom, going from about 80 to 57 percent of students. There are also fewer students who receive subsidized lunch or special education services. The largest change is in academics. Average baseline third grade scores are more than 0.7σ higher in AWC classrooms.

There are also differences between the AWC teaching corps and other teachers, as shown in columns 6 through 10. The causal effect of enrolling in fourth grade AWC is a statistically significant decrease in the proportion of novice teachers by 7 percentage points. Teachers have slightly more years of experience and are 9 percentage points less likely to have 5 or fewer years of experience, though these differences are statistically significant.

To investigate differences in teacher quality, I use a leave-out estimator of value-added to avoid bias from using value added as an outcome for students who directly contribute to the value-added estimate.[40] Value-added estimates are generated for the full state and standardized to be mean 0, standard deviation 1 so that the coefficient on value added represents the increase in value-added in terms of the standard deviation of Massachusetts teachers. The coefficients on value-added are small and positive—0.06 for math, on 0.15 for ELA—but not significant. These coefficients imply that AWC teachers have, on average, value added about a one-tenth of a standard deviation higher in the state distribution of teachers. However, this may not translate into impacts on college attendance. Chetty, Friedman, and Rockoff (2014b) (CFR) estimates that a one standard deviation increase in teacher quality

---

[39] Findings on MCAS outcomes for all years of data, the only outcomes available for all years of data, are in online Appendix Table B.8 and are similar to the main MCAS results.

[40] Despite using leave-out estimators of value added, the value-added estimates may still be biased by sorting on unobservables. If AWC teachers systematically have students sorted to them across years on dimensions not included in the control variables, the positive but not significant association between AWC and value added may be picking up this sorting rather than true differences in value-added. Estimating the value added of AWC teachers *not* teaching AWC students might account for this potential bias, but most teachers of AWC do not teach other classrooms or non-AWC classes in different years. Thus, I cannot estimate out-of-sample estimates of value added, and the estimates that I do use may be contaminated by sorting on unobservables.

TABLE 6—FUZZY REGRESSION DISCONTINUITY ESTIMATES OF EFFECTS ON FOURTH THROUGH SIXTH GRADE
CLASSROOM CHARACTERISTICS

	Black and Latino (1)	Subsidized lunch (2)	English learner (3)	Special education (4)	3rd grade MCAS (5)
Panel A. Peers					
2SLS	−0.233	−0.149	−0.140	−0.059	0.739
	(0.046)	(0.040)	(0.035)	(0.015)	(0.079)
CCM	0.804	0.846	0.397	0.122	−0.488
Observations (students)	6,900	6,900	6,900	6,900	6,900

	Math value-added (6)	ELA value-added (7)	Years experience (8)	Novice (9)	<5 years experience (10)
Panel B. Teachers					
2SLS	0.059	0.148	0.343	−0.066	−0.086
	(0.144)	(0.182)	(1.163)	(0.026)	(0.057)
CCM	0.246	0.345	11.238	0.128	0.309
Observations (students)	6,184	6,259	6,737	6,900	6,900

	MCAS index (11)	Class rank school (12)	Class rank classroom (13)	Days attended (14)	Days suspended (15)
Panel C. Outcomes					
2SLS	0.186	−2.522	−25.228	2.222	0.065
	(0.113)	(3.632)	(3.828)	(2.057)	(0.085)
CCM	0.147	68.451	71.062	171.431	−0.031
Observations (students)	6,859	6,859	6,858	6,899	4,502

Notes: Each coefficient labeled 2SLS is the fuzzy regression discontinuity estimate of fourth grade AWC attendance on the outcome listed in the column heading. An indicator for scoring above the AWC qualification threshold is the instrument for AWC attendance. The specification uses local linear regression with a triangular kernel of bandwidth 0.65. Listed below each coefficient is the CCM. All regressions include third grade year fixed effects. The sample is restricted to third graders enrolled in BPS in the fall of 2007 to 2012 in the grade levels that student-teacher-class links are available (fourth grade classrooms for the 2009–2012 third grade cohorts, fifth grade classrooms from the 2008–2011 third grade cohorts, and sixth grade classrooms for the 2007–2010 third grade cohorts). Third grade MCAS is the average of math and ELA scores. Regressions stack fourth, fifth, and sixth grade outcomes, include grade fixed effects, and triple cluster standard errors by baseline school by year, classroom, and student.

Source: Author's calculations from BPS and DESE data

increases college enrollment by 0.82 percentage points. Here, there is an average of a 0.1 standard deviation increase. Even if I take the upper bound on the confidence interval for value-added estimates (about 0.7 standard deviations in teacher value added), and multiply that by 2.2 (the average number of years spent in the program), that would only imply an increase of 1.3 percentage points in college attendance, using the CFR result as a benchmark. This is a small fraction of the AWC impact on college enrollment.

I also confirm in columns 11 through 15 that results for MCAS outcomes are similar between the main analysis sample and this more recent sample, but it is too soon to examine the more recent cohorts for longer-term outcomes. Here, AWC enrollment increases MCAS scores by about 0.19σ, and this finding is marginally statistically significant. This is slightly larger than the MCAS effects in the older sample.

In the more recent years of data, I can estimate class rank within school and within classroom. This shows that while there is no change in class rank at the school level, within the AWC classroom there is a significant decrease in class rank, which is to be expected with marginal students entering a classroom of high-achieving peers.

B. *Correlations between Potential Mechanisms and AWC Effects*

AWC is an amalgamation of several program components: the specialized curriculum, the particular school the AWC program is located in, the change in peer characteristics, and the designated AWC teachers. Some of these characteristics will vary by AWC classroom. To investigate if changes in classroom experiences are what is driving the AWC effects, I estimate individual sending-school reduced form effects of the AWC offer on MCAS scores and the reduced form effect on several potential channels, including take-up of the AWC offer (the first stage), peer characteristics, teacher value added, attendance in the AWC years, and suspensions in the AWC years.[41] Not all AWC placements have the same change in these characteristics. Thus if peers, for example, are a main mediator of the AWC effect, I would expect sending schools that have large impacts on MCAS to also have corresponding impacts on peers (or another characteristic).

A more formal approach would be to use site-specific interactions with the offer as instruments for several potential channels serving as multiple endogenous variables, but that would require meeting several additional assumptions beyond the standard fuzzy regression discontinuity setup.[42] Since the motivation for the graphs follows the intuition of the multi-site IV, I discuss the necessary assumptions here, which are detailed in Reardon and Raudenbush (2013). First, there must be constant treatment effects. If sending-school treatment effects are heterogeneous and covary with sending-school take-up rates, estimates using a multiple-IV strategy would be biased.[43] The concern in this context would be, for example, if there are larger effects for sending schools with a larger share of black and Latino students and the availability of the AWC program (and thus the take-up rate) covaries with those treatment effects. Second, potential channels would need to vary *only* as a consequence of receiving AWC (the exclusion restriction). AWC is a package of services that will directly vary based on taking up the AWC offer, but there may also be indirect channels. For example, AWC programs may be located in schools with particularly effective administrators or enthusiastic parents, and these factors may influence outcomes not directly through AWC. (Alternatively, one could consider the AWC treatment the bundle of both direct and indirect components.) Since both of these assumptions may not hold in practice, this exploration of mechanisms should be considered suggestive.

---

[41] Given the small sample sizes at some sending schools, I do not use fuzzy regression discontinuity estimates, as they can generate quite large outliers in schools where compliance with the AWC offer is low. Sending-school regressions must include at least 50 students to be included in the figures and correlations.

[42] For examples of the multi-site instrumental variables approach, see Kling, Liebman, and Katz (2007) for implementation in an RCT and Taylor (2014) for implementation in an FRD.

[43] Several recent papers have explored ways to address this concern, including Kirkeboen, Leuven, and Mogstad (2016); Kline and Walters (2016); Hull (2018); Mountjoy (2018).
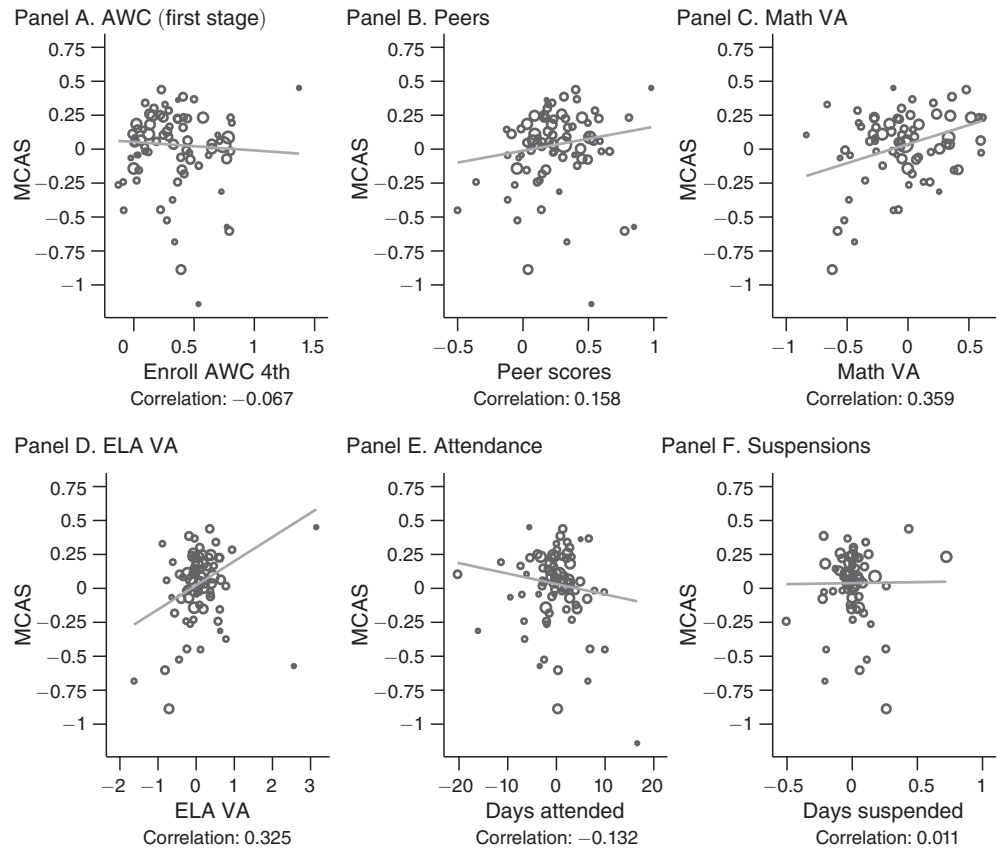
FIGURE 7. SCHOOL-LEVEL AWC OFFER EFFECTS ON CHANNELS BY AWC OFFER EFFECTS ON MCAS

Notes: This figure plots school-level estimates of scoring above the AWC eligibility threshold on MCAS against corresponding estimates on potential channels, including enrolling in AWC, baseline peer MCAS scores, math value-added (VA), ELA VA, attendance, and suspension. School-level estimates are weighted by number of students, with larger school samples represented by larger circles. The estimates are generated using the same regression discontinuity framework as described elsewhere in this study, using intent-to-treat estimates for each school and a larger bandwidth of 1, given the smaller amount of data. At least 50 students must be present in the regression for the school to be included in the plots and correlations. A single outlier with a large MCAS effect and small sample size is excluded from the plots but contributes to the correlations.

Source: Author's calculations from BPS and DESE data

Figure 7 plots these sending-school-level potential mechanism effects against the sending-school-level MCAS effects and reports the correlation of the AWC eligibility effect on MCAS and the corresponding mechanism effect, weighted by the number of students in each school. Interestingly, there is no relationship between MCAS offer effects and take-up of the AWC offer. It does not appear that the sites where more students attend the AWC program have larger impacts on test scores. Despite the large jump in peer scores, the correlation between peer baseline scores and MCAS effects is 0.158. This means that the students coming from sending schools where they experience the largest increases in peer scores tend to have slightly larger increases in test scores, but the relationship is weak and not statistically significant. There is a statistically significant relationship for math and

ELA value-added, with correlations with MCAS effects of 0.36 and 0.33 respectively. This means that the largest gains in AWC programs are coming from sending schools where there are also gains in value-added. But, as discussed above, while the differences in value-added are associated with test score gains, they likely are not large enough to account for more than a small proportion of the college effects observed in this study. There is no relationship between days attended or suspended and MCAS impacts. Overall, these relationships show little role for either take-up of the AWC program or peer effects to be driving MCAS gains, and some role for teachers. In short, it appears that AWC slightly increases the likelihood that a student has a high value added teacher and that these teachers increase test scores.

A lack of classroom links in earlier years of data prevent a similar exercise with teacher quality or other classroom characteristics in the older cohorts of data that have college outcomes. However, Figure 8 shows similar graphs for potential paths that explain the college effect, again using sending-school-specific reduced form estimates. Potential channels include take-up of the AWC offer (the first stage), elementary MCAS scores, elementary school attendance, elementary school suspensions, peer scores—though now peer scores are the baseline ELA scores of a student's peers at the school-level (rather than the classroom-level)—exam school attendance, SAT test taking, on-time twelfth grade progression, and on-time high school graduation.

A contrast with the classroom sample is immediately clear: sending schools that have higher AWC enrollment also have larger impacts on college enrollment ($\rho = 0.34$).[44] There are positive relationships between all of the sending-school-level effects on potential pathways and sending-school-level effects on college enrollment, except for MCAS scores ($\rho = -0.18$) and suspensions ($\rho = -0.12$), where there are negative but not significant relationships. AWC eligibility effects on elementary MCAS have little relationship with effects on college. This stands in contrast to Boston charter schools, where the schools with the largest MCAS gains also had the largest SAT and college impacts (Angrist et al. 2016), and is perhaps an indicator that the gains from AWC run through noncognitive channels (Cunha et al. 2006). Attendance ($\rho = 0.14$), peer baseline scores from potential AWC years ($\rho = 0.13$), and ever attending an exam school ($\rho = 0.12$) offer effects all have positive but small and nonsignificant correlations with college effects. The correlation between SAT-taking effects and on-time enrollment effects is significant and of similar magnitudes to teacher value added in the classroom sample ($\rho = 0.36$).[45] There are also strong relationships between arriving on time in twelfth grade and graduating high school on time and on-time college enrollment ($\rho_{ontime12th} = 0.31$, $\rho_{graduate} = 0.47$). In general, students from sending schools that have large impacts on college effects also tend to have big effects on milestones

---

[44] Note that this does not necessarily violate the constant treatment effects assumption discussed above since the offer estimate has not been rescaled by the first stage. Doing this rescaling results in no correlation with the take-up rate (not shown), but there is variation in sending-school-specific treatment effects. Given the small sample sizes, it is difficult to detect if this is due to sampling variation or a pattern in the type of sending schools with larger treatment effects.

[45] The correlation for AP test taking is slightly smaller than the coefficient on SAT test taking and is excluded from the plot for space reasons.
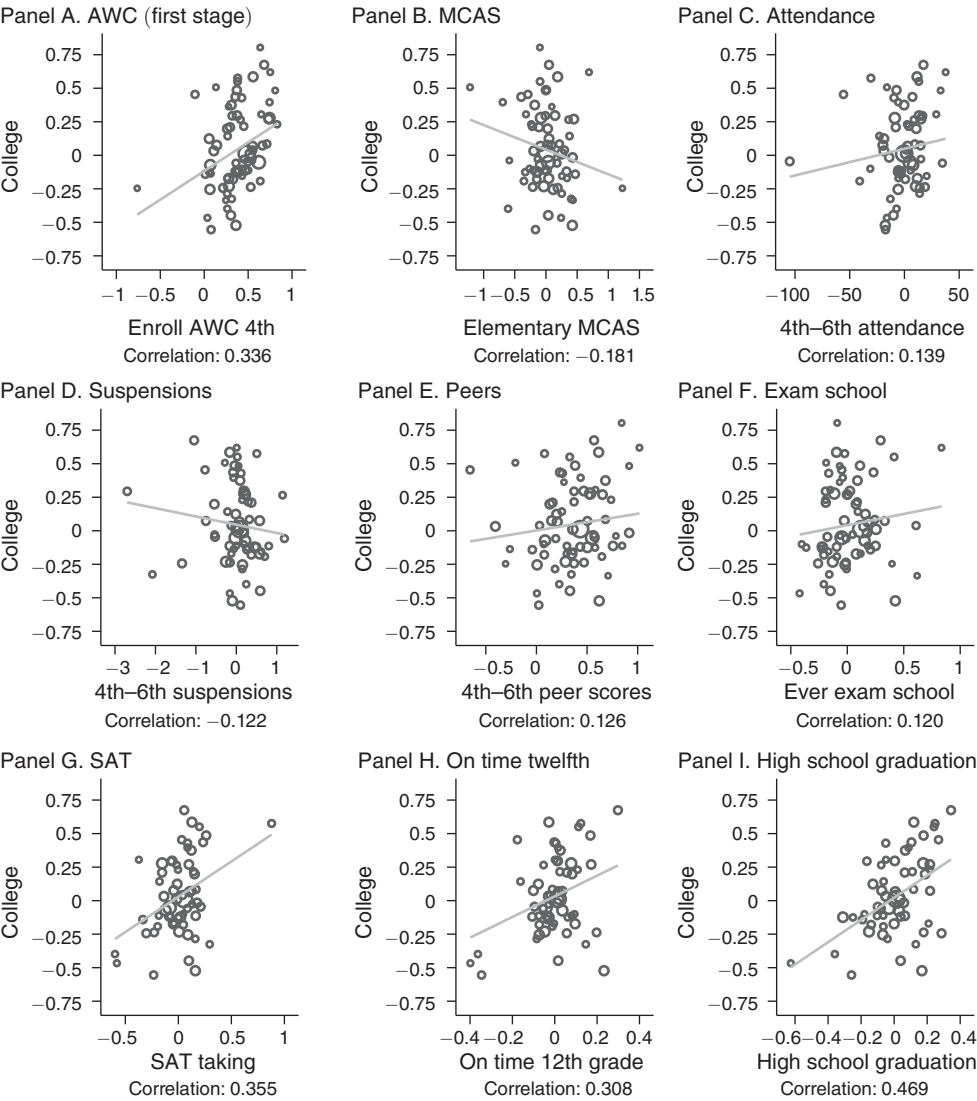
FIGURE 8. SCHOOL-LEVEL AWC OFFER EFFECTS ON CHANNELS BY AWC OFFER EFFECTS ON COLLEGE

Notes: This figure plots school-level estimates of scoring above the AWC eligibility threshold on college enrollment against corresponding estimates on other outcomes, including enrolling in AWC, MCAS, baseline peer MCAS scores, elementary school attendance, elementary school suspensions, attending an exam school, taking the SAT, and graduating high school. School-level estimates are weighted by number of students, with larger school samples represented by larger circles. The estimates are generated using the same regression discontinuity framework as described elsewhere in this study, using intent-to-treat estimates for each school and a larger bandwidth of 1, given the smaller amount of data. At least 50 students must be present in the regression for the school to be included in the plots and correlations.

Source: Author's calculations from BPS, DESE, and NSC data

that are necessary for college enrollment. Inasmuch as AWC increases students' likelihood of meeting these precursors, it also increases the likelihood that they attend college, providing further evidence of the importance of remaining on track in high school for eventual college enrollment.

VII. Conclusion

This paper suggests that a specialized program for high-achieving students can increase the long-term performance of students. AWC has positive but imprecise effects on standardized test scores. It increases on-time high school graduation. Perhaps most importantly, AWC has a large effect on college enrollment. These gains are primarily driven by black and Latino students, indicating that under-represented students in particular may benefit from interventions like AWC. The program does not, however, increase enrollment in BPS, nor does it affect exam school outcomes, two of the goals of the program. I also show that the fuzzy regression discontinuity approach behind these causal effects is generally robust to a number of specifications. An analysis of mechanisms suggests that the AWC effect on staying "on track" is particularly important and that meeting milestones like on-time progress through high school, taking the SAT, and on-time high school graduation are crucial for the pipeline to college enrollment. Peer effects appear to play little role in AWC gains.

The findings also highlight how typical it is for students to fall "off track" and that relatively high-achieving students are also susceptible. In particular, black and Latino students who just miss qualifying for AWC perform at the BPS average, despite almost qualifying for an elite program. AWC insulates black and Latino students from this trajectory, such that their high school graduation and college enrollment outcomes are on par with those of white and Asian students who participate in the program. Given the returns to college education (Hoekstra 2009; Carneiro, Heckman, and Vytlacil 2011; Oreopoulos and Petronijevic 2013; Zimmerman 2014), the large impact on college attendance for black and Latino students suggests that interventions like AWC can change the life courses of these students.

REFERENCES

Abadie, Alberto. 2002. "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variables Models." *Journal of the American Statistical Association* 97 (457): 284–92.

Abadie, Alberto. 2003. "Semiparametric Instrumental Variable Estimation of Treatment Response Models." *Journal of Econometrics* 113 (2): 231–63.

Abdulkadiroğlu, Atila, Joshua Angrist, and Parag Pathak. 2014. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." *Econometrica* 82 (1): 137–96.

Abdulkadiroğlu, Atila, Parag A. Pathak, and Christopher R. Walters. 2018. "Free to Choose: Can School Choice Reduce Student Achievement?" *American Economic Journal: Applied Economics* 10 (1): 175–206.

Allensworth, Elaine, Takako Nomi, Nicholas Montgomery, and Valerie E. Lee. 2009. "College Preparatory Curriculum for All: Academic Consequences of Requiring Algebra and English I for Ninth Graders in Chicago." *Educational Evaluation and Policy Analysis* 31 (4): 367–91.

Angrist, Joshua D., Sarah R. Cohodes, Susan M. Dynarski, Parag A. Pathak, and Christopher R. Walters. 2016. "Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice." *Journal of Labor Economics* 34 (2): 275–318.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444–55.

Bailey, Drew, Greg J. Duncan, Candice L. Odgers, and Winnie Yu. 2017. "Persistence and Fadeout in the Impacts of Child and Adolescent Interventions." *Journal of Research on Educational Effectiveness* 10 (1): 7–39.

Barrow, Lisa, Diane Whitmore Schanzenbach, and Amy Claessens. 2015. "The Impact of Chicago's Small High School Initiative." *Journal of Urban Economics* 87: 100–113.

Bettinger, Eric, Sten Ludvigsen, Mari Rege, Ingeborg F. Solli, and David Yeager. 2018. "Increasing Perseverance in Math: Evidence from a Field Experiment in Norway." *Journal of Economic Behavior and Organization* 146: 1–15.

Bui, Sa A., Steven G. Craig, and Scott A. Imberman. 2014. "Is Gifted Education a Bright Idea? Assessing the Impact of Gifted and Talented Programs on Students." *American Economic Journal: Economic Policy* 6 (3): 30–62.

Burke, Mary A., and Tim R. Sass. 2013. "Classroom Peer Effects and Student Achievement." *Journal of Labor Economics* 31 (1): 51–82.

Bursztyn, Leonardo, and Robert Jensen. 2015. "How Does Peer Pressure Affect Educational Investments?" *Quarterly Journal of Economics* 130 (3): 1329–67.

Calonico, Sebastian, Matias D. Cattaneo, and Max H. Farrell. 2018. "Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs." https://arxiv.org/pdf/1809.00236.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocio Titiunik. 2017. "Rdrobust: Software for Regression Discontinuity Designs." *Stata Journal* 17 (2): 372–404.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6): 2295–2326.

Card, David, and Laura Giuliano. 2014. "Does Gifted Education Work? For Which Students?" NBER Working Paper 20453.

Card, David, and Laura Giuliano. 2016. "Can Tracking Raise the Test Scores of High-Ability Minority Students?" *American Economic Review* 106 (10): 2783–2816.

Carneiro, Pedro, James J. Heckman, and Edward J. Vytlacil. 2011. "Estimating Marginal Returns to Education." *American Economic Review* 101 (6): 2754–81.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. 2017. "Simple Local Polynomial Density Estimators." https://eml.berkeley.edu/~mjansson/Papers/CattaneoJanssonMa_LocPolDensity.pdf.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma. 2018. "Manipulation Testing Based on Density Discontinuity." *Stata Journal* 18 (1): 234–61.

Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star." *Quarterly Journal of Economics* 126 (4): 1593–1660.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014a. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value added Estimates." *American Economic Review* 104 (9): 2593–2632.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014b. "Measuring the Impacts of Teachers II: Teacher Value added and Student Outcomes in Adulthood." *American Economic Review* 104 (9): 2633–79.

Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2017. "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility." NBER Working Paper 23618.

Clark, Damon. 2010. "Selective Schools and Academic Achievement." *B.E. Journal of Economic Analysis and Policy* 10 (1): 1–40.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2007. "How and Why Do Teacher Credentials Matter for Student Achievement?" NBER Working Paper 12828.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2015. "The Aftermath of Accelerating Algebra: Evidence from a District Policy Initiatives." *Journal of Human Resources* 50 (1): 159–88 .

Cohodes, Sarah R., and Joshua S. Goodman. 2014. "Merit Aid, College Quality, and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy." *American Economic Journal: Applied Economics* 6 (4): 251–85.

Cook, Philip J., Kenneth Dodge, George Farkas, Roland G. Fryer, Jr., Jonathan Guryan, Jens Ludwig, Susan Mayer, Harold Pollack, and Laurence Steinberg. 2014. "The (Surprising) Efficacy of Academic and Behavioral Intervention with Disadvantaged Youth: Results from a Randomized Experiment in Chicago." NBER Working Paper 19862.

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov. 2006. "Chapter 12—Interpreting the Evidence on Life Cycle Skill Formation." In *Handbook of the Economics of Education*, Vol. 1, edited by Eric A. Hanushek and Finis Welch, 697–812. Amsterdam: North-Holland.

Davis, Billie, John Engberg, Dennis Epple, Holger Sieg, and Ron Zimmer. 2013. "Bounding the Impact of a Gifted Program on Student Retention Using a Modified Regression Discontinuity Design." *Annals of Economics and Statistics* (111/112): 10–34.

Dee, Thomas S., and Emily K. Penner. 2017. "The Causal Effects of Cultural Relevance: Evidence from an Ethnic Studies Curriculum." *American Educational Research Journal* 54 (1): 127–66.

Dobbie, Will, and Roland G. Fryer, Jr. 2014. "The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools." *American Economic Journal: Applied Economics* 6 (3): 58–75.

Dobbie, Will, and Roland G. Fryer, Jr. 2015. "The Medium-Term Impacts of High-Achieving Charter Schools." *Journal of Political Economy* 123 (5): 985–1037.

Dougherty, Shaun M. 2018. "The Effect of Career and Technical Education on Human Capital Accumulation: Causal Evidence from Massachusetts." *Education Finance and Policy* 13 (2): 119–48.

Dougherty, Shaun M., Joshua S. Goodman, Darryl V. Hill, Erica G. Litke, and Lindsay C. Page. 2017. "Objective Course Placement and College Readiness: Evidence from Targeted Middle School Math Acceleration." *Economics of Education Review* 58: 141–61.

Elsner, Benjamin, and Ingo E. Isphording. 2017. "A Big Fish in a Small Pond: Ability Rank and Human Capital Investment." *Journal of Labor Economics* 35 (3): 787–828.

Eren, Ozkan, Briggs Depew, and Stephen Barnes. 2017. "Test-Based Promotion Policies, Dropping Out, and Juvenile Crime." *Journal of Public Economics* 153: 9–31.

Figlio, David N., and Marianne E. Page. 2002. "School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Inequality?" *Journal of Urban Economics* 51 (3): 497–514.

Gelman, Andrew, and Guido Imbens. 2019. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics* 37 (3): 447-56.

Goyer, J. Parker, Julio Garcia, Valerie Purdie-Vaughns, Kevin R. Binning, Jonathan E. Cook, Stephanie L. Reeves, Nancy Apfel, Suzanne Taborsky-Barba, David K. Sherman, and Geoffrey L. Cohen. 2017. "Self-Affirmation Facilitates Minority Middle Schoolers' Progress along College Trajectories." *Proceedings of the National Academy of Sciences* 114 (29): 7594–99.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica* 69 (1): 201–09.

Harris, Doug N., and Tim R. Sass. 2011. "Teacher Training, Teacher Quality and Student Achievement." *Journal of Public Economics* 95 (7–8): 798–812.

Hess, Frederick M. 2014. "America's Future Depends on Gifted Students." *New York Times*, June 4. https://www.nytimes.com/roomfordebate/2014/06/03/are-new-york-citys-gifted-classrooms-useful-or-harmful/americas-future-depends-on-gifted-students.

Hoekstra, Mark. 2009. "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach." *Review of Economics and Statistics* 91 (4): 717–24.

Hull, Peter. 2018. "IsoLATEing: Identifying Counterfactual-Specific Treatment Effects with Cross-Stratum Comparisons." http://www.mit.edu/~hull/isoLATE_022018.pdf.

Hurwitz, Michael, Jonathan Smith, Sunny Niu, and Jessica Howell. 2015. "The Maine Question: How Is 4-Year College Enrollment Affected by Mandatory College Entrance Exams?" *Educational Evaluation and Policy Analysis* 37 (1): 138–59.

Hyman, Joshua. 2017. "ACT for All: The Effect of Mandatory College Entrance Exams on Postsecondary Attainment and Choice." *Education Finance and Policy* 12 (3): 281–311.

Imbens, Guido W., and Joshua D. Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62 (2): 467–75.

Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies* 79 (3): 933–59.

Imberman, Scott A., Adriana D. Kugler, and Bruce I. Sacerdote. 2012. "Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees." *American Economic Review* 102 (5): 2048–82.

Jackson, C. Kirabo. 2010. "Do Students Benefit from Attending Better Schools? Evidence from Rule-Based Student Assignments in Trinidad and Tobago." *Economic Journal* 120 (549): 1399–1429.

Jackson, C. Kirabo. 2014. "Do College-Preparatory Programs Improve Long-Term Outcomes?" *Economic Inquiry* 52 (1): 72–99.

Jackson, Kirabo, and Alexey Makarin. 2018. "Can Online Off-the-Shelf Lessons Improve Student Outcomes? Evidence from a Field Experiment." *American Economic Journal: Economic Policy* 10 (3): 226–54.

Kane, Thomas J., and Douglas O. Staiger. 2008. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." NBER Working Paper 14607.

Katz, Lawrence F., Jeffrey R. Kling, and Jeffrey B. Liebman. 2001. "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment." *Quarterly Journal of Economics* 116 (2): 607–54.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad. 2016. "Field of Study, Earnings, and Self-Selection." *Quarterly Journal of Economics* 131 (3): 1057–1111.

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

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75 (1): 83–119.

Lavecchia, A.M., H. Liu, and P. Oreopoulos. 2016. "Chapter 1—Behavioral Economics of Education: Progress and Possibilities." In *Handbook of the Economics of Education*, Vol. 5, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 1–74. Amsterdam: North-Holland.

Lavy, Victor, M. Daniele Paserman, and Analia Schlosser. 2012. "Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom." *Economic Journal* 122 (559): 208–37.

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2): 281–355.

Mariano, Louis T., Paco Martorell, and Tiffany Tsai. 2018. "The Effects of Grade Retention on High School Outcomes: Evidence from New York City Schools." https://www.rand.org/content/dam/rand/pubs/working_papers/WR1200/WR1259/RAND_WR1259.pdf.

Marsh, Herbert W. 1987. "The Big-Fish-Little-Pond Effect on Academic Self-Concept." *Journal of Educational Psychology* 79 (3): 280–95.

Mcclain, Mary-Catherine, and Steven Pfeiffer. 2012. "Identification of Gifted Students in the United States Today: A Look at State Definitions, Policies, and Practices." *Journal of Applied School Psychology* 28 (1): 59–88.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2): 698–714.

Mountjoy, Jack. 2018. "Community Colleges and Upward Mobility." https://economics.stanford.edu/sites/g/files/sbiybj9386/f/mountjoy_2018_community_colleges_and_upward_mobility.pdf.

Murphy, Richard, and Felix Weinhardt. 2018. "Top of the Class: The Importance of Ordinal Rank." NBER Working Paper 24958.

Oreopoulos, Philip, Robert S. Brown, and Adam M. Lavecchia. 2017. "Pathways to Education: An Integrated Approach to Helping At-Risk High School Students." *Journal of Political Economy* 125 (4): 947–84.

Oreopoulos, Philip, and Uros Petronijevic. 2013. "Making College Worth It: A Review of the Returns to Higher Education." *Future of Children* 23 (1): 41–65.

Papay, John P., and Matthew A. Kraft. 2015. "Productivity Returns to Experience in the Teacher Labor Market: Methodological Challenges and New Evidence on Long-Term Career Improvement." *Journal of Public Economics* 130: 105–19.

Papay, John P., Richard J. Murnane, and John B. Willett. 2010. "The Consequences of High School Exit Examinations for Low-Performing Urban Students: Evidence from Massachusetts." *Educational Evaluation and Policy Analysis* 32 (1): 5–23.

Petrilli, Michael J. 2011. "Unrealized Potential." *New York Times*, October 2. https://www.nytimes.com/roomfordebate/2011/10/02/are-top-students-getting-short-shrift/unrealized-potential.

Pop-Eleches, Cristian, and Miguel Urquiola. 2013. "Going to a Better School: Effects and Behavioral Responses." *American Economic Review* 103 (4): 1289–1324.

Reardon, Sean F., and Stephen W. Raudenbush. 2013. "Under What Assumptions Do Site-by-Treatment Instruments Identify Average Causal Effects?" *Sociological Methods and Research* 42 (2): 143–63.

Rockoff, Jonah E. 2004. "The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data." *American Economic Review* 94 (2): 247–52.

Roda, Allison. 2015. *Inequality in Gifted and Talented Programs—Parental Choices about Status, School Opportunity, and Second-Generation Segregation*. Basingstoke, UK: Palgrave Macmillan.

Sacerdote, Bruce. 2011. "Chapter 4—Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" In *Handbook of the Economics of Education*, Vol. 3, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 249–77. Amsterdam: North-Holland.

Silverman, B.W. 1986. *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Steele, Claude M., and Joshua Aronson. 1995. "Stereotype Threat and the Intellectual Test Performance of African Americans." *Journal of Personality and Social Psychology* 69 (5): 797–811.

Taylor, Eric. 2014. "Spending More of the School Day in Math Class: Evidence from a Regression Discontinuity in Middle School." *Journal of Public Economics* 117: 162–81.

Walters, Christopher R. 2018. "The Demand for Effective Charter Schools." *Journal of Political Economy* 136 (6): 2179–2223.

Zimmerman, Seth D. 2014. "The Returns to College Admission for Academically Marginal Students." *Journal of Labor Economics* 32 (4): 711–54.