The Long-Run Impacts of Tracking High-Achieving Students: Evidence from Boston's Advanced Work Class*

Sarah Cohodes
Harvard University
cohodes@fas.harvard.edu

Draft:
Please do not circulate or cite without permission

April 2014

Abstract

Previous work on tracking high-achieving elementary and middle students in the US has shown little impact on short-run test scores. I provide the first estimates of the long-run impacts of tracking using data from the Boston Public Schools' (BPS) program for high-achieving students, Advanced Work Class (AWC). AWC is an accelerated curriculum in 4th through 6th grades with dedicated classrooms. BPS offers AWC to students who score well on a 3rd grade exam. Using a fuzzy regression discontinuity approach, I estimate the causal effect of AWC on standardized test scores, AP, SAT, high school graduation and college entrance. Like other programs for high-achieving students, AWC has little impact on test scores. However, it improves longer-term academic outcomes. AWC increases Algebra 1 enrollment by 8th grade, AP exam taking, especially in calculus, and college enrollment. It also has large positive effects on high school graduation for minority students. College enrollment increases are particularly large for elite institutions. One year of AWC attendance triples the rate of matriculation at a "most competitive" university. Using a multiple instrument strategy, I test several potential channels for program effects to operate and find suggestive evidence that teacher effectiveness and math acceleration account for AWC effects, with little evidence that peer effects contribute to gains.

*For making the data available for this project, I thank Kamal Chavda of the Boston Public Schools and Carrie Conaway of the Massachusetts Department of Elementary and Secondary Education and their staffs. Special thanks go to Josh Angrist, Chris Avery, and Larry Katz for their guidance and feedback. I am grateful to Natalie Bau, Josh Goodman, Peter Hull, László Sándor, and seminar participants at Harvard for helpful comments and discussions. Financial support from the Rappaport Institute for Greater Boston, the Taubman Center for State and Local Government, and the Multidisciplinary Program on Inequality and Social Policy, all at the Harvard Kennedy School, and institutional support from the School Effectivness and Inequality Institute at MIT are gratefully acknowledged.

1 Introduction

Tracking in schools – the practice of separating students into classrooms by ability – is hotly debated in the United States. Advocates for tracking claim that it helps teachers target instruction and ensures that higher-ability children have the opportunity to reach their maximum potential (Petrilli, 2011; Hess, 2014). Opponents claim that tracking places low-income and minority students in watered-down classes that exacerbate existing inequalities (Oakes, 2005). The evidence of tracking effect on student achievement is mixed (Betts and Shkolnik, 2000; Figlio and Page, 2002) and it is difficult to to isolate the effect of tracking from other endogenous inputs to the the educational production function.[1] A few recent studies take advantage of natural experiments or field trials to carefully isolate the effect of tracking. In an experiment that randomly assigned tracking to over 100 schools in Kenya, Duflo et al. (2011) find that tracking benefits both high- and low-achieving students, with high-achieving students benefiting through a positive peer effect and low-achieving students benefiting from targeted instruction despite the low-achieving peer context. Evidence from a policy in Chicago that designates students for extra instructional time in algebra based on test scores shows that students tracked into classrooms with low-ability peers have higher academic performance, though here the tracking effect is coupled with increased time on subject and support for classroom teachers (Cortes and Goodman, 2014).

Two common methods of tracking in the US are specialized instruction for students that are labeled "gifted and talented" and magnet schools for high achievers, often with entrance to the programs based on some form of testing. There is little well-identified research on gifted and talented programs at the elementary and middle school level, with two major exceptions. Bui, Craig, and Imberman (2014) study gifted and talented programs in a large urban school district utilizing both school lotteries and regression discontinuities. They do not find evidence of significant program impacts on test scores except for science scores, despite documenting a large change in peer characteristics. Card and Giuliano (2014) study a different large school district using a regression discontinuity approach and find few test score impacts for students identified as gifted by an IQ

[1] See Betts (2011) for an overview of the difficulties in estimating the effect of tracking, as well as a literature review of various approaches.

test. There are some gains in writing scores for those who qualify under a lower IQ threshold due to being from an underrepresented group, and gains in math, reading, and science for students who qualify for the program based on achievement tests rather than IQ tests. Research on magnet high schools also shows little effect on student achievement. Abdulkadiroğlu et al. (2014) and Dobbie and Fryer (2014) use regression discontinuities to estimate the effect of attending an magnet school with test-based admissions criteria in 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.[2]

Prior work on tracking for high-achieving students at the elementary and middle school level is limited by a short time horizon. A long-established program that tracks high-achieving students in the Boston Public Schools (BPS) provides the first opportunity to study the longer-term effects of this type of program for younger students. Advanced Work Class (AWC) is an accelerated program in the BPS for 4th through 6th graders who score well on a 3rd grade standardized test.[3] Students in the AWC program get a dedicated classroom with high-achieving peers, advanced literacy curricula, and accelerated math in the later grades. Since admission to the program is based on the 3rd grade test score, I compare students who scored just above and just below the admissions threshold to form causal fuzzy regression discontinuity estimates of the effect of the program on student outcomes. The long time horizon of the AWC program allows me to not only estimate the impact of AWC on state standardized exams, but also to determine its effect on Advanced Placement (AP) course taking and scores, SAT taking and scores, high school graduation and college enrollment. Previous work on other programs for high achievers in elementary and middle school has found little effect on test scores and has not been able to assess the impact these programs have on other

[2]Studies of exam schools outside of the US tend to find more positive results. See Clark (2010) for evidence from the UK, Jackson (2010) for Trinidad and Tobago, and Pop-Eleches and Urquiola (2013) for Romania.

[3]BPS does not explictly label AWC a "gifted and talented" program, whereas the programs studied in Bui et al. (2014) and Card and Giuliano (2014) 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 70th percentile in each subject. In the district studied by Bui et al., students can meet program requirements in several ways, but one of them includes scoring above the nationally-normed 80th 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 Guiliano, 6 percent are identified as gifted and 13 percent are enrolled in gifted classrooms. Within 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.

outcomes.

This paper differs from the other papers on tracking for high-achieving US students in elementary and middle schools in three main ways, in addition to the local context. First, I have additional outcomes for students that allow me to assess the longer-run impact of the program using measures more directly related to human capital accumulation than scores on standardized tests, providing the first evidence on the longer-term effects of tracking for high-achieving students. Second, I have data on the full universe of public school students in Massachusetts, so that attrition is less of a concern in my setting. Third, with detailed information on classroom and teacher characteristics and multiple instruments, I can investigate the channels through which the AWC program operates.

Like previous papers that examine tracking for high-achieving students, I find that AWC has few short-term test score effects. As time goes on, however, the AWC effect appears in increased Algebra 1 enrollment by 8th grade and increased AP test-taking, with half of the gains coming from enrollment in AP Calculus. There is a large, positive impact on four-year high school graduation for minority students. AWC also increases college enrollment. In particular, AWC increases enrollment at elite institutions by 4 percentage points per year of AWC attendance. This gain in matriculation at "most competitive" institutions more than triples the rate of attendance for comparison students with one year of AWC enrollment. Using a multiple instrument strategy that takes advantage of the school-specific context of AWC, I test the extent to which three potential channels – peer quality, as measured by baseline test scores, teacher value-added, and a catch-all term for remaining program effects – account for AWC impacts on test scores. Suggestive evidence from this approach finds little scope for peer effects, with teacher effects a much more likely mechanism for the transmission of AWC effects. A similar analysis for college outcomes (which cannot include teacher or classroom characteristics because of data limitations) suggests that math acceleration is the most likely channel for the gains in enrollment at elite institutions.

This paper proceeds as follows. The next section details the AWC program and admissions policies. In Section 3, I describe the data and sample and in Section 4 my empirical strategy. I report results in Section 5 and discuss potential threats to validity in Section 6. Section 7 includes a discussion of potential channels for the AWC effect and Section 8 concludes.

2 Advanced Work Class

The Advanced Work Class program has been a part of BPS since before the Judge Garrity school desegregation decision in 1974.[4] 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 which supports the program across schools.[5] All AWC programs include common elements in English/language arts (ELA) and math. In ELA, the curriculum 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" which ask students to write responses to the material they have covered. In mathematics, 4th 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 5th and 6th grades, so that students cover additional material. The goal is for students to be prepared to take calculus in their senior year of high school, which entails pre-algebra in 7th grade and algebra in 8th 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.

Students are accepted into the program by their score on a nationally-normed standardized exam offered in the fall of 3rd grade. All 3rd grade students are tested, with an alternative exam offered for Spanish-speaking students.[6][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 the scores of the 3rd grade. In the 3rd grade cohorts from 2001 to 2012, the top 11 to 17 percent of the 3rd grade test-takers are offered the program, with more students becoming eligible as additional school AWC programs were put in place.[8]

[4]The allocation of AWC was part the school desegregation plan in Boston, and AWC seats were allocated with racial preferences, as were exam school seats in addition to the more widely known busing policy.

[5]I thank Ailis Kiernan of the BPS AWC curriculum office for describing the program to me.

[6]There are two citywide AWC programs for Spanish-speaking students.

[7]Boston residents who do not attend BPS schools are also offered the opportunity to take the exam.

[8]Notably, while these are the top achievers in BPS, the nationally-normed percentile rank equivalent of the threshold is around the 70th percentile in both math and reading. Since the threshold incorporates both math and

Importantly, not every BPS school that serves 3rd 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 may then choose an AWC program as part of BPS's school choice process. 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 5th grade if they attended in 4th grade, though students must make academic progress in AWC. In 5th grade, all students, including those already attending AWC, are retested and 6th 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 6th 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, since the BPS school choice process typically takes place only before school entry grade levels. Thus, another reason for the somewhat low take-up rate of AWC for those above the threshold 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. Years of AWC enrollment (Panel A) is a function of distance from the qualification threshold, with a jump in years of enrollment at the threshold of about three-quarters of a year. Students who score under the threshold do have an increase in enrollment in the program, up to about half a year of attendance. This is mostly due to students who qualify for 6th grade AWC on their 5th grade test, but also to due to a small number of appeals by families and teachers for students who just miss the cutoff. This can be determined by Panel B, which shows enrollment in 4th grade AWC. Very few students beneath the threshold enroll in the program immediately if they are not eligible according to the cutoff score. In Panel B, there is about a 40 percentage point jump in immediate enrollment. There is less than perfect compliance with the offer of enrollment since many families choose not to enroll if it involves switching schools. As described in detail later, I employ a fuzzy regression discontinuity empirical strategy to estimate program effects that account for imperfect compliance to the threshold rule – both for students who do not choose to enroll and for students who enroll despite not receiving an

reading, the combined national percentile is likely a little higher, but still well below the highest national achievement levels.

offer in 3rd grade.

3 Data and Descriptive Statistics

3.1 Data

The Boston Public Schools (BPS) provided records of all 3rd 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. Each 3rd grade cohort provides the basis for the sample that I follow over time. I include all students who took the 3rd grade exam, including students who repeated 3rd grade, which allows me to identify the AWC cutoff amongst the entire distribution of 3rd grade test takers.[9] I match these students to records from BPS that show student enrollment in AWC by year and grade level.

BPS calculates eligibility as follows. The 3rd grade math and reading raw scores are standardized to be mean zero and standard deviation one, with missing scores substituted for zeroes. 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, with more seats offered in more recent years. Students who take the Spanish language exam may qualify under either exam. I reconstruct the BPS eligibility process in my data, and test each possible combined score to see how it predicts enrollment in 4th grade AWC. I select as a given year's threshold the score that had the biggest first stage F-statistic.[10] Visual evidence from these thresholds in Figure 1 shows a discontinuous jump in years of AWC enrollment of about

[9]This means that students can be in the sample in multiple years. In practice, this happens very rarely, as grade repeaters are typically not near the threshold for AWC qualification so they are not in the sample limited by a bandwidth near the threshold. The restriction to BPS students at baseline means I exclude a small number of students who are enrolled in private schools but choose to take the test to see if they qualify for AWC, although these students are included in the calculation of distance to the threshold.

[10]BPS provided their official cutoff scores for a subset of years. The empirically derived thresholds are quite similar to the BPS thresholds in the years it is possible to compare to the two, but not exactly the same, likely due to minor differences in data. Since I do not have the official cutoffs scores for the earliest years of the sample (third grade cohorts from 2001 and 2002), I use the empirically determined cutoff scores for my analysis to be consistent across years and enable me to use the oldest cohorts, which are the only cohorts with available college outcomes. I include in my robustness checks results using the official cutoff (where possible) and find similar results using this specification.

three-quarters of a year of enrollment at these empirically derived thresholds, and similarly an increase of almost 40 percentage points in terms of 4th grade AWC enrollment.

The Massachusetts Department of Elementary and Secondary Education (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. I linked 3rd grade students to the Student Information System (SIMS) records to obtain demographic characteristics, baseline programmatic status as a special education student, English language learner, or subsidized lunch recipient. I also linked students to their 3rd grade Massachusetts Comprehensive Assessment System (MCAS) scores, as an alternative measure of student achievement from the Stanford 9 or TerraNova exam used to determine AWC eligibility.[11] 3rd grade ELA MCAS scores are available for all cohorts, and 3rd grade math MCAS scores are available since 2006.[12] I have access to the full universe of Massachusetts public school students, so I follow students throughout their academic careers even if they leave BPS, as long as they remain in Massachusetts public schools.

For school years 2010-11 to 2013-14, DESE also provided Student Course Schedule (SCS) and Education Personnel Information Management System (EPIMS) records. These data allow me to link students and teachers to specific classrooms and courses. I use them to calculate classroom peer characteristics and teacher characteristics, including teacher value-added, for 4th through 6th grade classrooms in the available years. Peer characteristics are calculated using baseline (third grade) demographic, program participation, and test score information, grouped by the course identified in the student-teacher-course link. I calculate teacher valued-added using a specification with lagged tests scores, lagged score squares, and cubics, demographics, and peer demographics and tests following Kane and Staiger (2008). I use a leave-year-out estimator to reduce bias, as indicated in Chetty, Friedman, and Rockoff (2014a; 2014b), though this means I can associate

[11] Since MCAS exams are administered in the spring after students and their families are notified of AWC eligibility, it's possible that being above the threshold for AWC acceptance has an effect on 3rd 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, 3rd grade MCAS scores are not discontinuous at the threshold.

[12] The No Child Left Behind Act (NCLB) requires 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, Masschusetts had some exams in all grades 3 through 8 and 10, but in not all subjects.

a slightly smaller number of classrooms with teacher value-added than I can with other teacher characteristics. I calculate value-added estimates for 4th through 6th grade in ELA and math. I also use the SCS data to calculate enrollment in math courses by a particular grade level, e.g. Algebra 1 by 8th grade. The math class enrollment outcomes allow me to test whether AWC achieves its goal of math acceleration. I use the most common advanced math track in BPS, which is: 7th grade, prealgebra; 8th grade, algebra 1; 9th grade algebra 2; 10th grade, geometry; 11th grade, precalculus; and 12th grade calculus.[13] This is difficult to do in other subjects, as there is not a clear hierarchy of classes or an advanced track.

For outcomes, I connect the records of 3rd 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. I detail the specifics of each outcome below. Some outcomes are based on projected senior year in high school. I determine this by adding 10 to the fall year of 3rd grade. Unless otherwise specified, all outcome data comes from DESE.

- *Enrollment*: I track enrollment in 4th through 12th grade at any BPS school, a BPS exam school (a district 7th-12th grade magnet school with acceptance determined by test), Boston charter schools, and non-Boston Massachusetts public schools (including non-Boston charters). I separate enrollment in non-Boston Massachusetts public schools between those who enroll through METCO, a program that allows BPS students to register at suburban schools, and those who enroll through moving town of residence. These outcomes are all unconditional, so that students who leave the data (Massachusetts public schools) are counted as zeroes for the enrollment outcomes.

- *MCAS*: MCAS raw scores are standardized on the entire state population to be mean zero and standard deviation one. In grades 4 through 8 and 10, all students are tested in math and ELA in most years. Fourth, 7th, and 10th grade also include a writing exam. In all grade levels that writing is tested, it is scored on two dimensions: topic development and writing

[13]However, some students and schools deviate from this track: some students take geometry in 9th grade and algebra 2 in 10th grade. Students may also take a variety of courses in 11th grade, some of which are not labeled as precalculus.

composition (English grammar conventions). Science is included in 5th, 8th, and 10th grades. To increase precision, I stack elementary school (4th and 5th grade) and middle school (6th-8th grade) outcomes and double cluster the standard errors from relevant regressions by student and 3rd grade school.

- *Exam school application*: In addition to observing enrollment in an exam school, I observe application and offer data at exam schools, including scores on the ISEE, the test used for exam school admission.[14] Application and offer variables are unconditional. Unlike the test for AWC, student must choose to take the exam school entrance test. I observe exam school application for the fall 2001-2005 3rd grade cohorts.

- *AP and SAT*: AP and SAT are observed for the cohorts of 3rd graders who are in 12th grade in projected senior years of 2011 through 2014 for AP scores and 2011 to 2013 for SAT scores. 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 unconditional.

- *High school graduation*: I observe high school graduation from any school in Massachusetts for projected senior years of 2011 through 2014. I observe 5-year high school graduation for one fewer cohort. Again these, outcomes are unconditional.

- *College*: NSC data is available for 3rd graders with projected senior years of 2011 to 2013. I construct college enrollment measures from the NSC on college type (2- or 4- year, public or private, and Barron's selectivity ranking) within a within 6 months of time since expected high school graduation. All college outcomes are unconditional, with zeroes attributed to those who leave the sample. Notably, the NSC match for the first two college cohorts includes all students who were 8th graders in Massachusetts and some additional nongraduates, including those who later leave the sample, so that the NSC outcomes include almost all students in the relevant 3rd grade cohorts.[15]

[14] The data for these outcomes are the same data used in Abdulkadiroğlu et al. (2014).

[15] In the regression discontinuity sample, all students in the 2001 cohort were sent to NSC for matching, 90 percent of students in the 2002 cohort were sent to NSC, and 79 percent of students in the 2003 cohort. Nongraduates from the 2003 cohort have yet to be matched to the NSC, and I anticipate receiving this match, as well as an additional cohort of NSC data, in March 2015.

- *Peer and teacher characteristics*: Classroom characteristics are available for 3rd grade cohorts from 2007 through 2012, for whom student-teacher-course links are observed. Peer characteristics include demographics, special education, English language learner, and subsidized lunch status, and test scores from 3rd grade, averaged at the classroom level. Teacher characteristics include value-added, years of experience, and novice status. Essentially all teachers in Massachusetts are licensed and considered highly qualified under NCLB, so I do not compare teachers on these dimensions.

- *Math course enrollment*: Math course enrollment are available for the cohorts and grades that link to course data. The 3rd grade cohorts included by grade level are: 7th grade, 2006-2009; 8th grade, 2005-2008; 9th grade, 2004-2007; 10th grade, 2003-2006; 11th grade, 2002-2005; and 12th grade, 2001-2004.

In order to follow a consistent sample of students throughout the paper, I focus on the 3rd grade cohorts from 2001 to 2003. These are the students for whom I observe college outcomes. Since student-teacher-course links are only available for more recent 3rd grade cohorts, I use more recent cohorts for analyses on peer and teacher characteristics, and a variety of cohorts that link to math course enrollment information by grade level. I also present estimates of my main findings using all available 3rd grade cohorts for each outcome in Appendix B.

3.2 Descriptive Statistics

I limit my main analysis sample to students enrolled in BPS in 3rd grade in 2001 through 2003 who take the Stanford 9 test, and describe students based on their 3rd grade pre-AWC enrollment characteristics. Third graders in BPS as a whole generally come from a disadvantaged background. As shown Column 1 in Table 1, most 3rd grade BPS students receive subsidized lunch (84%) and are nonwhite (88%). About 15 percent of all 3rd graders are English language learners and 19 percent are special education participants. Third grade test scores are well below the state average. Compared to the entire population, AWC participants are more advantaged. About 6 percent of 4th and 5th graders are enrolled in AWC, and 9 percent of 6th graders. Column 2 of Table 1 indicates that those who enroll in 4th grade AWC are more likely to be girls, less likely to be black or Hispanic, more likely to be white or Asian, and less likely to received subsidized lunch or be an English language learner. Unsurprisingly, very few AWC enrollees are also identified as receiving special education services. They score over half a standard deviation above the state mean on 3rd grade MCAS, and most students who enroll in 4th grade continue on in AWC in the subsequent years. Importantly, while this population is less disadvantaged than the BPS population as a whole, 68 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 3rd grade test scores 0.3 standard deviations (σ) 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 RD 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. As seen in Column 4, which shows the characteristics for students above the threshold and outside the RD bandwidth (the highest-achieving students), black and Hispanic students are less like to have 3rd grade scores that put them far above the eligibility threshold. Asian students, who make up 35 percent of the highest scoring group, account for all of the English language learners in the highest-achieving group. Appendix Table A.3 shows which student characteristics predict years of AWC enrollment, both above and below the threshold, not limited to the RD sample.[16] Asian students are the racial group most likely to enroll, if given an offer.[17] Underneath the threshold, "always-takers" are typically high-achieving white or Asian students. Together, these descriptive facts account for an RD sample that has many more black students than the enrolled in AWC sample.

In terms of outcomes, I show in Table 2 that AWC outpace their peers in BPS. For MCAS scores, Boston students typically score 0.25 to 0.65σ below the state mean, whereas AWC students score 0.35 to 0.72σ above the mean. AWC students are much more likely to take an AP test or the SAT and to graduate high school.[18] Finally, 64 percent of AWC students enroll in any college

[16]These regressions are descriptive and do not have a causal interpretation.

[17]For more on the characteristics of those above and below the threshold who do and do not take up the treatment, see Appendix Tables A.4 and A.5.

[18]Note that the high school graduation rates shown here are lower than published graduation rates for the district, since they are based off 3rd grade year and include students that leave the sample as zeroes.

within 6 months of expected high school graduation, including two-year institutions, whereas 33 percent of the district as a whole does.[19] Again, the RD sample in Column 3 is somewhere between all students and AWC enrollees, but closer to the AWC means. 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 to 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 will determine if any of these positive outcomes associated with AWC students can be causally attributed to the program.

4 Empirical Framework

As discussed above, a raw comparison of students who enroll in AWC with other BPS students would be misleading. AWC students are much 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 et al., 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 distribution, since students cannot precisely control their score on a test. Within a small window of points on an exam, students are in 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 holds in the case of

[19] The college outcomes also include students who leave the sample as zeroes.

AWC. Since the threshold changes yearly and students 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.[20] This proves to be the case empirically. As shown in Figure 2, the frequency of test scores moves smoothly through the threshold, with no jump in frequency of a particular test score around the cutoff. There is some evidence of a sawtooth pattern – this is due to the relatively small number of potential combined scores in a given year's data, a pattern seen more dramatically in Appendix Figure B.2 where the more recent 3rd grade cohorts tested with the TerraNova have even fewer available combined scores, due to the small number of raw score points available on that exam.

In a further check on the soundness of the regression discontinuity, I show that student background characteristics are smooth functions across the the threshold in Appendix Figure A.1 and confirmed with regressions in Appendix Table A.1. Another 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. Even though I do not require that students remain in the data to be included in most of my analyses, I still note that there is little differential attrition, as shown in Table A.2. At one grade level (6th grade), students who are offered AWC are less likely to leave the sample, with a marginally significant differential of 6 percentage points. I will discuss attrition in more detail in Section 6.2, including strategies to account for this one small difference. Importantly, there is no significant differential attrition in the upper grades or for students who are not sent to the NSC for matching for college outcomes.

The threshold is determined by a cut score for the combined math and reading scores, as described in Section 3.1. I create a measure of distance to the threshold, *Gap*, by subtracting the threshold from the combined score.[21] Figure 1 shows that adherence to the threshold rule is not perfect. A few students just below the threshold enter AWC, mostly through the 6th grade entrance

[20] 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). Since IQ scores have a subjective element, test administrators might give students scores just above the threshold in order to give them access to gifted programming, either consciously or unconsciously.

[21] *Gap* is measured in numbers that look quite similar to effect sizes, but since the combination of z-scores is not itself mean zero standard devation one, it is not actually in standardized units.

but a handful through the appeals process. 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 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.

Because the effect of AWC is likely to accumulate over time spent in the program and in order to address partial compliance, I model outcomes as a function of years enrolled in the AWC program.[22] For a student i in the 3rd grade in school s in school year t, I estimate a system of local linear regressions of the following form:

$$YearsAWC_{ist+k} = \alpha_0 + \alpha_1 Above_{ist} + \alpha_2 Gap_{ist} + \alpha_3 Gap_{ist} \times Above_{ist} + \lambda' X_i + \delta_{st} + \epsilon_{ist} \tag{1}$$

$$Y_{ist+k} = \beta_0 + \beta_1 Years\hat{AW}C_{ist+k} + \beta_2 Gap_{ist} + \beta_3 Gap_{ist} \times Above_{ist} + \theta' X_i + \mu_{st} + \eta_{ist} \tag{2}$$

where Gap_{ist} measures distance to the AWC eligibility threshold on the 3rd grade, $Above_{ist}$ is an indicator variable for being above the threshold in a given year, $YearsAWC_{ist+k}$ is a count variable for the number of years of AWC enrollment in the school year $t+k$ after 3rd grade with a maximum of three, X_i is a vector of 3rd grade characteristics (gender, race, special education, limited English proficiency, and subsidized lunch status), and Y_{ist+k} is an outcome interest in some year, $t+k$, subsequent to 3rd grade. The causal impact of AWC is represented by β_1 from the second stage regression, with program enrollment instrumented by program eligibility, $Above_{ist}$. I include 3rd grade school by year fixed effects, δ_{st} and μ_{st}, respectively, since available AWC seats will be specific to a particular school and year, and all students in the same school and year will face the same

[22] See Angrist and Imbens (1995) for details on 2SLS with variable dosage endogenous treatments.

choice set of AWC programs.

My preferred model estimates local linear regression with a triangular kernel in a bandwidth of 0.5 on either side of the program cutoff. I fully saturate the model with baseline demographic and program participation covariates to increase precision. The triangular kernel weights points near the threshold more heavily than those distant from the threshold. I estimate optimal bandwidths for each outcome according to the Imbens and Kalyanaraman (2012) procedure. For simplicity, I use a bandwidth of 0.5, which is the Imbens-Kalyaramanan optimal bandwidth for the first stage (rounded up). I later test the robustness of my findings to several additional bandwidths, including the IK bandwidth computed for each outcome, and specifications. Standard errors are clustered by 3rd grade school.

I report the reduced form and 2SLS estimates where space allows. The reduced form estimates are the difference in outcomes between those above and below the threshold without taking into account program enrollment, within the allotted bandwidth, weighting points nearest the threshold. The 2SLS estimates are the causal impacts of the program for compliers. Note that I do not specify a particular channel through which the program works for the 2SLS estimate to be the causal effect for my main results. It may be through the specialized curriculum, the designated teachers, the peer group, or another factor.[23] 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 generates a program effect. The CCM is not directly observable, because those beneath threshold who do not enroll in AWC are a mix of compliers and students who would never enroll in AWC even if eligible. I estimate the CCM by taking outcome mean in the below the threshold group, which consists of "never-takers" and compliers, to use the potential outcomes language of Angrist, Imbens, and Ruben (1996), and subtracting off the outcome mean of the never-takers in the above the threshold group, adjusted by the AWC dosage in each of those groups, with the same bandwidth and weights as described above.[24] This

[23]I examine some of these channels in Section 7.

[24]It is possible to estimate a treatment complier mean in a similar manner. In this case, the "TCM" is the mean

is an adaptation of the measurement of the control complier mean in the context of a randomized experiment in Katz et al. (2001) to the fuzzy regression discontinuity setup using the methods discussed in Abadie (2002; 2003). Specifically, I estimate:

$$\begin{aligned} Y_{ist+k} * (1 - YearsAWC_{ist+k}) = \gamma_0 + \gamma_1 (1 - Years\hat{A}WC_{ist+k}) \\ + \gamma_2 Gap_{ist} + \gamma_3 Gap_{ist} \times Above_{ist} + \phi' X_i + \nu_{st} + \xi_{ist} \end{aligned} \tag{3}$$

where $1 - YearsAWC_{ist+k}$ is instrumented by AWC eligibility as in Equation 2 and γ_1 is the estimate of the control complier mean. I use the CCM as my measure of outcomes for the group beneath the threshold because alternative measures of the mean below the threshold will commingle outcomes for compliers with those of always-takers (if treated students are included) and never-takers (even if treated students are excluded) and thus be subject to selection bias.

5 Results

5.1 First stage and the effect on enrollment

First stage estimates of the years of AWC enrollment are in Table 3. The three columns account for the fact that AWC enrollment years vary based on the grade level of the outcome, with a maximum of one for 4th grade outcomes, two for 5th grade outcomes, and three for outcomes in 6th grade and later. For outcomes in 6th grade and above, the first stage effect of being above the AWC eligibility threshold is a 0.83 of a year jump in years of enrollment from around 0.44 years of enrollment for students just beneath the threshold.[25] Two factors contribute to this. First, there is jump in initial enrollment of 38 percentage points, as seen in Column 1. Then, of those who accept the AWC offer in 4th grade, on average, they stay in the program for about an additional 2.2 years ($\frac{0.83}{0.38}$) compared to those just below the threshold. Students below the threshold generally accumulate years of AWC enrollment by qualifying for the program in 6th grade. The first-stage F-statistic

of the treated group above the threshold, which consists of "always-takers" and compliers, with the mean for always-takers from the group underneath the threshold subtracted off, again adjusted for dosages and with the same default specification as previously described. It can be estimated in a manner similar to the one represented in Equation 3, using $YearsAWC_{ist+k}$ instead of $(1 - YearsAWC_{ist+k})$.

[25] In the first stage table, I report the mean of the first stage outcome for students within 0.05 units beneath the threshold instead of control complier means, since the CCM is not a meaningful concept for the first stage.

using years of AWC enrollment as the endogenous variable is 81.

As noted above, the initial increase in 4th grade AWC enrollment is a 38 percentage point increase in AWC enrollment. Fewer than 7 percent of students just below the threshold enroll in AWC when it is measured by 4th grade enrollment (Column 1), which is why I consider most noncompliance below the threshold to be due to 6th grade enrollment rather than the appeals process. For parsimony, in later results I do not repeat first stage estimates, which differ only slightly from the ones presented here based on the particular sample (for example, a few students are missing MCAS scores in a given grade). The first stage varies slightly by whether or not a school has an AWC program. Unsurprisingly, schools with AWC programs have larger first stages. I generate these first stage estimates by fully interacting the default specification with indicators for whether the 3rd grade school hosts an AWC program. Scoring above the threshold in a school that has an AWC program results in a first stage of 0.95 years of attendance (or 40 percentage points when using 4th grade AWC as the endogenous variable). At a school without an AWC program, the first stage is 0.79 years of attendance (or 37 percentage points of proportion enrolled in 4th grade AWC). Essentially, having an AWC program at a school induces about a 3 percentage point increase beyond that at a non-AWC school in initial enrollment, and this initial difference persists and magnifies over time.[26]

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 et al., 2014), AWC has few effects on the enrollment choices of students either during the grades that AWC serves or in subsequent grades, as shown in Appendix Table A.6. AWC does not influence enrollment at Boston exam schools, which are three magnet schools for high-achievers that also admit students based on test

[26]Appendix Table A.13 presents results by 3rd grade school characteristics. Panel A shows results seperately by 3rd graders in schools that have an AWC program in 4th grade and those that do not. There are few significant differences by school type, though as a whole it appears that students coming from schools with an AWC program score higher on the MCAS and have larger college effects, but students coming from schools without AWC have a larger AP Calculus effect. I will discuss these results in more details in Section 6.3.

scores. This may be because a large majority of students are applying to an exam school anyway, as shown in Appendix Table A.7.[27] 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 which go to BPS students, at least for students on the margin.

5.2 Achievement Outcomes

Like recent evaluations of gifted and talented programs (Bui et al., 2014; Card and Giuliano, 2014), AWC has little immediate effect on elementary school standardized test scores, as seen in Columns (1) of Table 4. To increase precision for the MCAS estimates, I stack elementary (4th and 5th) and middle school (6th through 8th) grades and double cluster the standard errors by student and 3rd grade school. Years of AWC enrollment is the endogenous variable, which means that 4th grade outcomes have a maximum of 1 for the endogenous variable, 5th grade outcomes 2, and 6th grade and higher outcomes, 3. I report reduced form and 2SLS outcomes – which illustrate how there are different possible dosages at each grade level. For elementary school outcomes, the reduced form is about half the size of the 2SLS, since the second stage estimate is scaled by a first stage estimate around 0.5 years (halfway between the 4th grade and 5th grade first stages reported in Table 3). For middle school and high school outcomes, the reduced form and 2SLS outcomes are very similar, since the first stage of 0.85 years is close to one. I also combine test score outcomes into one academic 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 subject are in Appendix Table A.8.

[27] The interaction between AWC enrollment and exam school application may have some explanatory power for the generally null results found in Abdulkadiroğlu et al. (2014). Seventy-one percent of students who enroll in AWC for at least one year apply to an exam school, with 82 percent of those who applied receiving an offer. About 36 percent of exam school applicants have attended at least one year of AWC, and about 58 percent of exam school offers go to those who have enrolled in AWC. If one thinks of AWC and exam school enrollment as essentially the same treatment, 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 already been treated. Indeed the one high school in Boston that shows some impacts on achievement outcomes in the regression discontinuity set up is the O'Bryant, which has the lowest proportion of AWC-treated students in the sample near the relevant exam school threshold. Another potential explanation is that there are interaction effects with age, with elementary and middle school treatment being more important than upper middle school and high school treatment. 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.

There are no significant impacts on the academic index for elementary or middle school students. The magnitudes are small positives and differ little for low-income or minority students. In 10th grade MCAS, there are also no significant differences, though the magnitude of the 2SLS AWC effect on the MCAS academic index is slightly larger at 0.07 σ per year of AWC attendance. The test score effect is particularly large for minority students, at 0.14σ per AWC year, though again this result is not statistically significant. MCAS is one of the few outcomes for which I have several additional cohorts of data, and the MCAS results change little when I use all available years of data, though the 10th grade score gains become marginally statistically significant. (Appendix Table B.6). One reason why there might be few impacts on test scores is that the high-achieving students who make up the RD sample are "topping-out" on the MCAS, i.e. scoring the very top score with no room to gain. This is not the case. Very few students in the RD sample score at the very top of the exam, and there is no differential effect on top scoring by AWC participation (results available by request).

If what matters for academic achievement is relative position in the academic distribution, as posited by Marsh (1987) (the "big-fish-little-pond-effect"), an investigation of whether or not AWC influences class rank is also relevant. Thus, I also show the effects of AWC on class rank within a school in Columns (4)-(6) of Table 4. I generate class rank by determining the percentile of a student's academic index in the distribution of scores in their school in that year and grade.[28] Class rank is measured between the 0th and 99th percentile, with larger numbers indicating the higher end of the score distribution. In the elementary years, AWC decreases school rank percentile, though this difference is not significant. This is likely due to the concentration of high-achieving students at a school with an AWC program. In middle school there is a small positive difference in school rank percentile, and in high school there is an increase in rank of 2.7 percentiles per year of AWC enrollment. Notably, compared to the control complier mean, the increase in high school rank essentially maintains the overall class rank to around the 65th percentile for those that attend AWC for 3 years, rather than increasing it. It is possible that the lack of change in class rank is what explains the lack of test score effects.

[28] I can do this procedure by classroom only for the more recent years of data, as shown later in Table 9.

Standardized test scores only tell a partial story in terms of academic potential. One of the main goals of the AWC program is to accelerate mathematics instruction. In Table 5 I examine whether or not AWC achieves this goal by estimating the effect of AWC on enrolling in a specific math course by a certain grade level. The typical advanced sequence in BPS is 7th grade prealgebra, 8th grade algebra 1, 9th grade algebra 2, 10th grade geometry, 11th grade precalculus, and 12th grade calculus. However, some schools switch the order of algebra 2 and geometry, and some offer a variety of 11th grade courses that are not explicitly labeled precalculus. Since course enrollment information is only available from DESE from school year 2010-2011 to school year 2013-2014, each outcome in Table 5 is measured for different cohorts. For example, the 3rd grade cohorts from fall 2005-2008 can be observed in 8th grade in the course enrollment data. Given this data limitation, I choose to show course outcomes for all available cohorts rather than limiting to the main analysis sample (cohorts from 2001-2003).

Algebra 1 is a precursor for college mathematics, and there are policy movements to increase algebra 1 enrollment at earlier grades Panel (2008). However, the evidence on the impact of algebra is mixed. Studies using nationally representative samples find a positive association between algebra and education and other outcomes, but are subject to selection bias (Stein et al., 2011; Rickles, 2013). Policies instituting universal algebra for 8th or 9th graders can have adverse effects (Allensworth et al., 2009; Clotfelter et al., 2012a), because students who are not academically prepared for algebra must also enroll. But effects are heterogenous; universal policies can have beneficial effects for high-achieving students (Clotfelter et al., 2012b). Given that AWC-eligible students are at the higher end of the achievement distribution, enrollment in algebra by 8th grade is likely to be beneficial. As can be seen in Column (2) of Table 5, there is a large, significant increase in enrollment in algebra 1 by 8th grade, of 12 percentage points per year of AWC attendance. With a control complier enrollment rate of 60 percent, this implies that essentially all students who attend AWC for 3 years will enroll in algebra 1 by 8th grade. However, there is not a corresponding bump in 7th grade prealgebra enrollment. This is likely due to the lack of specific labeling of 7th grade math courses in the course enrollment data. Similarly, there is no corresponding significant effect on enrollment in the advanced math track through 9th to 12th grades in high school, although there is a positive coefficient of around 2 to 7 percentage points per year of attendance at each course by grade outcome. The lack of effect on the high school grades may be due to inconsistent labeling in the course data or a variety of potential course sequences that all lead to calculus in 12th grade, or it may be a lack of effect after 8th grade, or different effects by cohorts. Thus, the gains in algebra 1 by 8th grade are suggestive evidence that AWC is successful in accelerating mathematics, at least in middle school. More years of course data are needed to determine if there is an effect on other grades. As I will discuss later, there is gain in AP Calculus taking, suggesting that part of the math acceleration effect is a switch from regular calculus to the AP offering.

In Table 6, I present estimates for key high school outcomes that are related to success in higher education and in general: AP, SAT, 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. AWC participants are more likely than their counterparts to take an AP exam, with a significant 9 percentage point increase exam participation per year of AWC. About half of the overall increase in AP exam taking is driven by a marginally significant increase of 4.6 percentage points in AP Calculus taking per year of attendance.[29] This means that one year of AWC attendance almost doubles the rate of AP Calculus taking. This finding is consistent with the small positive calculus increase in Table 5, where calculus enrollment includes non-AP Calculus, and also indicates that most of the increase in calculus enrollment is coming from the AP option or from switching to the AP track. The AP results also give the opportunity to examine not just course taking, but student achievement. A score of 3 on an AP exam is considered "qualified" for college credit. However, there are no effects on test scores for overall AP tests, or when considered by each subject. One of the goals of the AWC program is to prepare students to take calculus by their senior year of high school by accelerating the math curricula in 5th and 6th grade, and the results for AP Calculus taking and scores indicate that the program is indeed able to influence this outcome down the line.

Taking the SAT is another key milestone for application to college, as many four year colleges require the exam.[30] As seen in Table 6, control complier students take the SAT at the rate of

[29] For additional subject-specific AP results, see Appendix Table A.9.
[30] Colleges also accept the ACT, but most students in Massachusetts take the SAT.

72 percent, and AWC does not have a significant impact on SAT test taking or scoring above the Massachusetts median score.[31] AWC has a positive but not significant effect on high school graduation overall (using 3rd grade cohort year to calculate projected senior year), but gives a large boost to on-time high school graduation for minority students, with a gain in graduation rate of 12.8 percentage points per year of AWC attendance. Using the estimate on 5 year high school graduation of 6.8 percentage points, about half this increase is from a reduction in completion time and about half is from high school graduation that would not happen in absence of the program.

5.3 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 7, though this increase is not significant. This table shows college enrollment the fall after projected high school graduation, with projected high school graduation year calculated by adding 10 to the 3rd grade cohort year. Results (available by request) showing enrollment two falls after graduation are very similar. Sixty percent control compliers enroll on time, and there is a gain of 5.7 percentage points per year of enrollment for AWC participants, though this effect is not significant. This enrollment effect comes from increased matriculation at both four- and two-year institutions. Within four-year institutions, AWC shifts enrollment from public universities to private universities.

The question of whether AWC enrollment shifts college type beyond sector is also relevant. Arguably causal evidence on the quality of a higher education indicates that attending a higher quality institution can increase graduation rates (Cohodes and Goodman, 2014) and earnings (Hoekstra, 2009). I measure college quality through enrollment at a highly selective university, as categorized by Barron's rankings.[32] There is a large, statistically significant effect on on-time enrollment in a "most competitive" college of 4.2 percentage points per year of AWC attendance. Very few control complier students enroll in these elite institutions; with 2 percent of these students

[31] See Appendix Table A.10 for subject specific results.

[32] "Most competitive" institutions include Tufts University and Boston College, the two most commonly attended highly selective institutions in my sample. It also includes the Ivy League schools and elite liberal arts colleges.

enrolling, the AWC effect more more than triples that enrollment rate with one year of AWC attendance. I show the reduced form relationship between distance from the threshold and college enrollment in Figure 3. The increase at the threshold for matriculation at most competitive is visually apparent in Panel B. The one-year magnitude of the effect on elite college attendance is similar to the one found in Deming et al. (2013), where attending a first-choice (higher-quality) school resulted in an increase in enrollment at selective institutions by 4.2 percentage points. However, when multiplied by the 3 potential years of AWC attendance, it is larger than the effect detected in Charlotte by Deming et al.. It stands in contrast to results on elite college-going for other educational interventions in Boston. Angrist et al. (forthcoming) find that attendance at a Boston charter school increases four-year college enrollment by about 18 percentage points – but they find no effect on attending highly selective institutions. Abdulkadiroğlu et al. (2014) find no effect of attending a Boston exam school on either overall enrollment or enrollment at elite institutions.

Enrollment effects are quite large for minority students. Black and Hispanic students are 10 percentage points more likely to enroll in college per year of AWC attendance, and the majority of this gain is from enrollment at four-year institutions. The switch to the private sector for four-year colleges is particularly large for minority students, with a 13 percentage point per year of AWC attendance increase in four-year private enrollment. This finding is significant at the 10 percent level. While the gains at the most elite institutions is of similar magnitude for minority students as for all students, no control complier minority students enroll at these elite institutions.[33] This low rate of elite matriculation among control compliers is consistent with given recent research documenting the phenomenon of "under-matching" among disadvantaged youth (Hoxby and Avery, 2012; Hoxby and Turner, 2013), and these results suggest that AWC counters the under-matching phenomenon.

[33]Since the control complier mean is an estimated result, it is technically possible to have CCM's that are negative, as seen in Panel C. However, since CCM's are estimated with some error, these very small negatives can be considered equivalent to zero.

6 Threats to Validity

6.1 Robustness

The results are robust to a number of specification checks. In Table 8, I present results for key outcomes for a variety of specifications and bandwidths, including the Imbens-Kalyaramanan ("IK") bandwidths and bias-corrected estimates and bandwidths from the procedure described in Calonico, Cattaneo, and Titunik (forthcoming) ("CCT"). Panel A replicates my default specification for reference purpose. Panel B varies the specification, first excluding the baseline covariates, then using the official BPS cutoffs where available – which limits results to the 2003 cohort alone, then excluding the 2001 cohort, and also using a quadratic functional form on the full sample. Panel B also reports the CCT estimates, which both select a bandwidth and adjust the estimates and standard errors for bias.[34] Panel C shows a larger bandwidth (0.75) and a smaller bandwidth (0.25). It also includes the optimal bandwidths from the IK procedure on the reduced form estimates of each outcome, which range between 0.45 and 1.27.

When I use my original specification but remove controls for demographics and 3rd grade program participation, there are few changes in the magnitude or significance of the effects, though the standard errors are slightly larger (as expected, since I fully saturated the default specification in order to increase power). The findings of an increase in enrollment at most competitive institutions remain statistically significant. As discussed in Section 3.1, I have official cutoff scores from BPS for the 2003 cohort (and other younger cohorts). When I substitute the BPS official cutoff in that one year, my results are generally similar. However, there are no longer any significant effects in the results for the cohort from 2003, likely because the sample size is cut by two-thirds. The finding on attending elite universities remains of similar magnitude, though there is a negative coefficient on on-time four-year enrollment. This is likely due to worse NSC coverage for the 2003 cohort, which will be remedied with an additional NSC match in Spring 2015. For the algebra 1 by 8th grade outcome, I can substitute the official cutoffs for all years of data contributing to

[34] The statistical package that accompanies the CCT procedure does not allow covariates, so these estimates do not include covariates or year by school fixed effects. Results generated by using the CCT bandwidth but otherwise using my default specification yield similar, though slightly smaller, results.

that outcome. Here, the results are substantively the same, with an even larger effect size using the official thresholds. I also estimate my findings excluding the 2001 cohort. As can be seen in Appendix Table A.12, college effects are particularly large for this cohort. Excluding 2001 leaves results that are similar, but of smaller magnitude and no longer significant. This is a cause for caution when viewing the results as a whole, which I will address with additional years of data as students age into 12th grade and college outcomes.

I also fit quadratic polynomials on either side of the threshold, using the whole sample and no weights. The parametric approach yields similar results, with the enrollment effects at elite institutions remaining and the high school MCAS results becoming significant.[35] Estimates using the CCT procedure tend to have much smaller bandwidths and larger coefficients. Comparing the CCT results to the estimates in Panel C for the bandwidth of 0.25 shows that part of this increase is due to the tightening of the bandwidth and part to the bias correction procedure. Since this is a new econometric technique, I consider the CCT results suggestive that the effect of AWC may be larger than the findings from my default model, but do not consider it conclusive evidence.

In Panel C, I vary the bandwidths but continue to use local linear regression with a triangular kernel with baseline controls. Generally, magnitudes are larger with the 0.25 unit bandwidth and slightly smaller with the 0.75 unit bandwidth. As the IK bandwidths for the most part are larger the default bandwidth of 0.5, results using optimal bandwidths also have somewhat smaller magnitudes, though they remain statistically significant and follow the same pattern as the main findings. My selection of the 0.5 point bandwidth has little effect on my conclusions, and throughout all of my robustness checks my general findings remain the same. Notably, the gains in on-time enrollment at elite institutions are of similar magnitudes in all of the robustness checks and statistically significant in most.

6.2 Attrition

As discussed in Section 4, there is little differential attrition by program eligibility, as shown in Appendix Table A.2. The exception is 6th grade, where students above the AWC cutoff are more

[35] Following Gelman and Imbens (2014) I do not estimate parametric models with higher order polynomials.

likely to leave the sample. In addition to this, in the high school grades, there is a somewhat high level of overall attrition, with around 20 percent of the control compliers not appearing in the data in 9th through 12th grades. These students either leave the state, attend private schools, or drop out of high school. The state sends almost all students in my sample to match to the NSC, my source for college information, as seen in Column (10).[36] To address the concern that the somewhat high level of attrition or the differential attrition in 6th grade might bias my findings, where possible, I rerun my analyses to account for attrition.

While the overall level of attrition in elementary MCAS outcomes is small, it reaches about 12 to 17 percent for control compliers in middle school and 22 percent for the control compliers in 10th grade, leaving room for the MCAS outcomes to be influenced by attrition. To address this possibility, in all grade levels, I substitute the baseline test score for missing test score outcomes. Since 3rd grade ELA scores are the only baseline scores available in the time period I am using, I use 3rd grade ELA scores to substitute for missing academic index outcomes (which are also on a standardized scale). I present the results of this substitution in Appendix Table A.11. There are very little differences between this table and Table 4. There is no consistent pattern of differences between the results excluding attriters and those where baseline scores are substituted for missing scores, and all effects remain not significant. For NSC outcomes, I have one cohort of students (in 3rd grade in the fall of 2001) who all were sent to the NSC for matching. When I rerun my college estimates on this subsample in Appendix Table A.12, results for college enrollment are even larger, despite the decrease in sample size. However, as discussed above, it is possible that the 2001 cohort is anomalous for reasons other than complete follow up in the NSC. Given the consistent findings from the MCAS and college analyses modified for attrition, my findings do not appear to be biased by the level of attrition.

[36]This is because DESE sends most nongraduates to the NSC who enroll in at least 8th grade in a Massachusetts high schools and has occasionally conducted additional matches for researchers. Currently, the 2003 cohort is missing the nongraduate match but the previous two cohorts are not. In the regression discontinuity sample, 100 percent of the 2001 3rd grade cohort has been sent to the NSC for matching, 90 percent of the 2002 cohort, and 79 percent of the 2003 cohort. An additional match in Spring 2015 will bring up the match rate for the 2003 cohort.

6.3 Contamination effects

In the context of a randomized controlled trial, contamination effects occur when some treatment other than the one being tested influences the control group, which could potentially account for the effects seen (or not seen) on the treatment group. In the fuzzy regression discontinuity framework for AWC, a contamination effect could explain the positive outcomes I find if something occurred that made student compliers below the AWC threshold *worse off* while those above the threshold remained at previous levels of achievement. The most likely candidate for contamination is the program itself: AWC removes high-achieving peers from the classrooms of students just below the threshold. If those students are providing a positive peer effect, AWC could make students below the threshold worse off. On the other hand, if AWC creates more homogenous classrooms and which allows teachers to better target their instruction, the removal of high-achieving peers could have beneficial effects, as found in (Duflo et al., 2011).

To test the concern that contamination effects are driving my results, I estimate effects by school-level AWC eligibility rate. First, I calculate the school level percentage of students eligible in a 3rd grade cohort in each year. This rate ranges between 0 percent and over 50 percent, with a median of 7.6 percent. Appendix Figure A.2 shows the distribution of school-level AWC eligibility (weighted by students) for all 3rd grade students (Panel A) and for the regression discontinuity sample (Panel B). I then divide the sample into two groups: those with below median school-level eligibility rates ("low eligibility") and those with above median school-level AWC eligibility rates ("high eligibility"). To estimate results by these groups, I fully interact the default specification used above with indicators for low and high eligibility. If contamination effects are driving my results, I would expect effects that I attribute to AWC to be larger for the high eligibility group, since these are the schools for which the peer composition will change most dramatically. As can be seen in Panel C of Appendix Table A.13, there are no significant differences between groups based on eligibility rates, and no consistent pattern of results. Students from high eligibility schools have higher initial test score effects (Columns 1 and 2), but lower high school test effects (Column 3). Algebra 1 (Column 4) and college gains (Columns 8 and 9) seem to be higher for students from low eligibility schools. And results for AP and high school graduation outcomes (Columns 5-7) appear substantively the same. If anything, on the longer term outcomes, it appears that the students with the least scope for contamination effects are those with the largest results.

7 Mechanisms

In the estimates above, I have not specified a specific channel through which the AWC program generates its effects. It could be some specific aspect of the program, or it could be that AWC set students on an accelerated track that later generates the college effects. This section will discuss potential mechanisms, first documenting that there is a difference in classroom experiences between AWC and non-AWC classrooms. In Table 9 AWC classrooms are different than the alternate classrooms attended by control compliers. These results for 4th through 6th 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. Specifically, they include 4th grade classrooms for the 2009-2012 3rd grade cohorts, 5th grade classrooms from the 2008-2011 3rd grade cohorts, and 6th grade classrooms for the 2007-2010 3rd grade cohorts – *not* the cohorts used in the main analysis sample above. However, I have no reason to believe that the AWC program differed in the first three cohorts from the more recent ones with classroom data available. Here, I use AWC attendance in 4th grade as the endogenous treatment rather than years of AWC, since it does not make sense to discuss classroom composition in terms of years of exposure. Panels A and B show that the classroom composition, as measured by demographic characteristics and other 3rd grade characteristics, is dramatically different based on AWC treatment. As first shown observationally in Table 1, the causal effect of AWC on classroom composition is fewer black and Hispanic students and more white and Asian students. There are fewer students who receive subsidized lunch or special education services. Baseline 3rd grade scores are substantially higher.

There are also statistically significant differences between the AWC teaching corps and other teachers, as shown in Panel C, again using 4th grade AWC as the endogenous variable. The causal effect of enrolling in 4th grade AWC is a decrease in proportion of novice teachers by 6 percentage points. However, on average, there is no difference in teacher years of experience.[37] Prior papers on

[37] There are also no differences by gender or race (not shown).

tracking programs for high-achievers do not have value-added estimates for teacher effectiveness, likely because of the data needed to calculate these effects. With the full state of Massachusetts data as well as student-teacher-class links, I can estimate value-added differences induced by the program. As noted above, 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, and I calculate value-added scores for each ELA and math. The coefficients on value-added are small and positive, but not significant.[38] I also confirm in Panel D that results for MCAS outcomes, returning to the use of years of AWC as the endogenous variable, 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. 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.

AWC is an amalgamation of several program components, some of them described above: the specialized curriculum, the particular school the AWC program is located in, the change in peer characteristics, and the designated AWC teachers. The first item on this list affects all AWC programs similarly, and thus it is difficult to tease out its influence on AWC treatment effects. The particular school that AWC students enroll in is endogenous, since it is influenced by already being enrolled in a school with AWC or which AWC programs a family chooses to list on their school choice form. However, the latter two aspects of the program will vary by AWC classroom, and I can adapt my fuzzy regression discontinuity framework to include those particular treatments with some additional assumptions and modifications of the empirical strategy.

As in Abdulkadiroğlu et al. (2014), I use the offer of AWC to instrument for multiple endogenous variables that describe the treatment – peer baseline test scores and teacher value-added. I also

[38]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.

include years of AWC exposure as an additional channel to describe all other aspects of the AWC treatment not explicitly identified through the peer or teacher channels. In order to identify multiple endogenous treatments, I need at least the same number of instruments as endogenous variables. To obtain sufficient instruments, I consider the AWC eligibility system a multi-site regression discontinuity, as in Taylor (2014). I create multiple instruments by interacting the offer variable with each 3rd grade school. While students at all schools face the same cutoff in a given year, the AWC offer varies by school, since some schools have AWC programs and some do not, so the AWC offer at each school will vary in practice by the availability of AWC in that school and other nearby schools. I then use these multiple school-offer variables in an over-identified 2SLS framework, with multiple endogenous variables. The intuition behind this approach is that the school-specific offer of AWC "randomizes" not only the AWC treatment within a small neighborhood around the threshold, but it also randomizes a bundle of school services. For example, a student under the threshold at a given school will get a particular combination of teachers, peers, and other inputs to the educational production function. And a student over the threshold will get a different combination of teachers, peers, other inputs, and AWC. Since not all AWC programs (or alternative placements) have the exact same bundle of services, the school-specific instruments can identify effects when there is variation in aspects of the treatment. The multiple endogenous variables analysis using classroom characteristics is limited to the recent cohorts.

I present results using the school-specific instruments in Table 10. Each column within a panel displays the results from a single regression with the school level instrument; Columns (3)-(7) use multiple endogenous variables. The outcome is the academic index. In Panel A, I use teacher value-added as a measure of teacher quality induced by the AWC offer. However, given the concern that value-added estimates will be biased by sorting on unobservables to AWC teachers, Panel B shows results from the same empirical setup, with novice teachers substituted for value-added. Given that on average, novice teachers have lower value-added than their more experienced counterparts (Rockoff, 2004), Panel B offers another way to assess the impact of teacher quality without the potentially biased value-added score. First, in Column 1 I estimate the effect on the academic index of years of AWC, instrumented with the multiple offers. As expected, the results here are very similar to the MCAS comparison results in Table 9. In Columns 2 and 3, I use the alternative endogenous variables – peer scores and teacher value-added/novice teacher– each separately in their own regression, instrumented by the multiple offers. Peer scores are the average classroom baseline 3rd grade MCAS math and ELA scores, and value-added is the standardized (on the full state) sum of math and ELA value-added. Novice teachers are represented by an indicator for having a teacher with 1 year of experience or less. The results for baseline peer quality indicate that an increase of one standard deviation in peer scores through the AWC program, would, on average, increase the academic index by about about 0.10σ, though this relationship is not statistically significant.

When value-added is used as the endogenous variable with multiple instruments, there is a large positive coefficient on value-added, indicating that an increase in one standard deviation in teacher quality, as measured by value-added, would increase the academic index by 0.24σ. As discussed above, this relationship may be biased by unobserved sorting to AWC teachers. In Panel B, when the AWC offer induces students to have a novice teacher, the effect on the academic index is almost a full negative standard deviation. The novice teacher endogenous variable only has a first stage F-statistic of 7.5, so this finding should only be considered suggestive. However, along with the significant positive effect on teacher value-added, the negative coefficient on novice teachers adds to the evidence that teachers are a very important channel for the transmission of AWC effects. In both cases, when teacher quality measures are combined with peer scores and/or years of AWC in the multiple endogenous variables 2SLS estimates shown in Columns 3 through 7, the teacher channel typically has the largest and most statistically significant effect on the academic index. In the case of novice teachers, when that variable is included with all other endogenous variables, it is no longer a weak instrument, and the coefficient remains a large, though not statistically significant negative.

In no cases is the peer score or years of AWC coefficient statistically significant, and in most cases the coefficients are quite small. The coefficient on peer scores ranges between about 0.05σ and 0.15σ, similar to the modest coefficients on peer effects found in much of the literature (see Sacerdote (2011) for an overview). As a whole, I take this evidence to mean that when all of the channels are considered together, years of AWC and peer effects are the least likely channels for transmission of AWC gains, a finding in line with many other recent explorations of peer effects in elite schools (Abdulkadiroğlu et al., 2014; Dobbie and Fryer, 2014; Bui et al., 2014). Changes in teacher quality induced by the offer of AWC seem a much more promising channel for how students accumulate AWC gains.

Due to data limitations, I cannot conduct a similar multiple endogenous variables analysis with teacher quality in the older cohorts of data that have college outcomes. However, I can conduct a similar exercise, again using school-specific offers, but using the different potential channels available in the data for the older cohorts. In this case, I continue to include years of AWC and peer scores – though now peer scores are the *school* 3rd grade ELA scores of the students in a particular school, averaged over all grade levels that a student is observed in the data. I also include AP course-taking, SAT taking, and on-time high school graduation as potential channels to be instrumented by the school-based offer variables. I separate AP taking into AP Calculus and all other APs, to examine if accelerated math has a particular impact on the college outcomes. I present the results of the multiple endogenous variables analysis in Table 11. Panel A uses on-time enrollment in 4 year institutions as the outcome, and Panel B on-time enrollment at a most competitive school.

For on-time four year college enrollment, each of the potential channels considered separately has a positive and significant effect on on-time 4-year enrollment. This 2SLS setup with each channel considered separately implies that the AWC effect transmits solely through each variable considered. This is not a plausible assumption, so considering all of the channels are jointly, as in Column (7), is a more realistic setup for how AWC might induce enrollment changes. Here, only SAT-taking and on-time college enrollment have significant effects. They imply that an AWC-induced change in SAT-taking or high school graduation behavior will have a large effect on on-time college enrollment, while years of AWC, peers, and APs do not not contribute. Since 4-year on-time college enrollment includes any 4-year institution, no matter the selectivity, the emphasis on SAT and high school graduation makes sense. The latter is probably a mechanical effect: it is impossible to enroll on-time in college without first graduating high school. The SAT effect is likely about switching students from nonselective institutions to those that require test scores.

The results for enrollment at most competitive institutions focus on other channels (Panel B).

Here, when all potential channels are considered together, only enrollment in AP Calculus induced by the AWC offer contributes to the elite matriculation effect. This implies that the elite enrollment effect is due to AWC's emphasis on math acceleration. As for test score outcomes at younger grades, for both college outcomes, there is little evidence that peer effects are a channel through which AWC operates. Instead, it appears that basic college preparation activities are important for on-time enrollment at a 4-year institution, with math acceleration particularly important for enrollment at an elite college.

8 Conclusion

This paper has shown that a tracking program for high-achieving students can have significant, positive effects on the long-term performance of students, despite having little impact on state standardized test scores. Instead, AWC increases Algebra 1 enrollment by 8th grade and AP course taking, particularly in AP Calculus; it also increases on-time high school graduation for minority students. Perhaps most importantly, AWC has a large effect on elite college enrollment. The program does not, however, increase enrollment in the Boston Public Schools nor does it affect exam school outcomes, two of the goals of the program. Some critics of tracking suggest that high achieving students will still do well in the absence of tracking or other specialized programs for them. The impacts of AWC on elite college attendance suggest that the trajectories of high-achieving students can be altered by their schooling experiences. Given the evidence that college quality can affect college graduation and earnings (Cohodes and Goodman, 2014; Hoekstra, 2009), this is a particularly important outcome. Other studies that do not have long time horizons would imply that similar programs have little impact. This paper shows that outcomes other than standardized test are important for showing gains for high-achieving students. Even in cases with short time horizons, it may be possible to study other important outcomes like mathematics acceleration.

I also show that the fuzzy regression discontinuity approach behind these causal effects is robust to a number of specifications and that the regression discontinuity setup is sound. Using a multiple instrument strategy, I test several potential channels for program effects to operate and find suggestive evidence that teacher effectiveness and accelerated mathematics are plausible mechanisms for the transmission of AWC effects. As with other studies of programs that group high-achieving students in the US, I find that peer effects have little influence on future outcomes (Abdulkadiroğlu et al., 2014; Dobbie and Fryer, 2014). This raises the question of whether a dedicated program like AWC is necessary to achieve similar results. It is possible that policies focusing directly on math acceleration for high-scoring students or teacher effectiveness outside of the AWC model would have similar beneficial effects for high-achieving students.

The findings from this analysis resonate with a number of studies of educational interventions that find initial short-term effects that fade out over time, only to resurge later in long-run outcomes (Chetty et al., 2011; Dynarski et al., 2013; Garces et al., 2002). However, in the case of AWC, there are no detectable short-term impacts. This could be due to insufficient outcome measures during the program, or because the program only affects outcomes by setting students on academic trajectories that later influence outcomes.

References

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variables models. *Journal of the American Statistical Association 97*(457), 284–292.

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics 113*(2), 231–263.

Abdulkadiroğlu, A., J. Angrist, and P. Pathak (2014). The elite illusion: Achievement effects at Boston and New York exam schools. *Econometrica 82*(1), 137–196.

Allensworth, E., T. Nomi, N. Montgomery, and V. 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), 367–391.

Angrist, J. D., S. R. Cohodes, S. M. Dynarski, P. A. Pathak, and C. R. Walters (forthcoming). Stand and deliver: Effects of Boston's charter high schools on college preparation, entry, and choice. *Journal of Labor Economics*.

Angrist, J. D. and G. W. Imbens (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association 90*(430), 431–442.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association 91*(434), 444.

Betts, J. R. (2011). The economics of tracking in education. In S. M. a. L. W. Eric A. Hanushek (Ed.), *Handbook of the Economics of Education*, Volume Volume 3, pp. 341–381. Elsevier.

Betts, J. R. and J. L. Shkolnik (2000). The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review 19*(1), 1–15.

Bui, S. A., S. G. Craig, and S. 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.

Caloncio, Sebastian, M. D. C. R. T. (Forthcoming). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*.

Card, D. and L. Giuliano (2014). Does gifted education work? For which students? Working Paper 20453, National Bureau of Economic Research.

Chetty, R., J. N. Friedman, N. Hilger, E. Saez, D. W. Schanzenbach, and D. Yagan (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *Quarterly Journal of Economics 126*(4), 1593–1660.

Chetty, R., J. N. Friedman, and J. E. Rockoff (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review 104*(9), 2593–2632.

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

Clark, D. (2010). Selective schools and academic achievement. *The B.E. Journal of Economic Analysis & Policy 10*(1), 1935–1682.

Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor (2012a). The aftermath of accelerating algebra: Evidence from a district policy initiative. Working Paper 18161, National Bureau of Economic Research.

Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor (2012b). Algebra for 8th graders: Evidence on its effects from 10 north carolina districts. Working Paper 18649, National Bureau of Economic Research.

Cohodes, S. R. and J. S. Goodman (2014). Merit aid, college quality, and college composition: Massachusett's Adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics 6*(4), 251–85.

Cortes, K. E. and J. S. Goodman (2014). Ability-tracking, instructional time, and better pedagogy: The effect of double-dose algebra on student achievement. *American Economic Review 104*(5), 400–405. WOS:000338925400068.

Davis, B., J. Engberg, D. Epple, H. Sieg, and R. Zimmer (2013). Bounding the impact of a gifted program on student retention using a modified regression discontinuity design. *Annals of Economics and Statistics* (111-113).

Deming, D., J. Hastings, T. Kane, and D. Staiger (2013). School choice, school quality and postsecondary attainment. *American Economic Review 104*(3), 991–1013.

Dobbie, W. and R. G. Fryer (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.

Duflo, E., P. Dupas, and M. Kremera (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *The American Economic Review 101*(5), 1739–1774.

Dynarski, S. M., J. M. Hyman, and D. W. Schanzenbach (2013). Experimental evidence on the effect of childhood investments on postsecondary degree attainment and degree completion. *Journal of Policy Analysis and Management 32*(4), 692–717.

Figlio, D. N. and M. E. Page (2002). School choice and the distributional effects of ability tracking: Does separation increase inequality? *Journal of Urban Economics 51*(3), 497–514.

Garces, E., D. Thomas, and J. Currie (2002). Longer-term effects of head start. *The American Economic Review 92*(4), 999–1012.

Gelman, A. and G. Imbens (2014). Why high-order polynomials should not be used in regression discontinuity designs. Working Paper 20405, National Bureau of Economic Research.

Hahn, J., P. Todd, and W. Van der Klaauw (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica 69*(1), 201–209.

Hess, F. M. (2014, June). America's future depends on gifted students. *The New York Times*.

Hoekstra, M. (2009). The effect of attending the flagship state university on earnings: A discontinuity-based approach. *The Review of Economics and Statistics 91*(4), 717–724.

Hoxby, C. M. and C. Avery (2012). The missing "one-offs": The hidden supply of high-achieving, low income students. Working Paper 18586, National Bureau of Economic Research.

Hoxby, C. M. and S. Turner (2013). Exapnding college opportunities for high-achieving, low income students. SIEPR DIscussion Paper 12-014, Stanford Institute for Economic Policy Research, Stanford, CA.

Imbens, G. and K. Kalyanaraman (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies 79*(3), 933–959.

Jackson, C. K. (2010). Do students benefit from attending better schools? evidence from rule-based student assignments in trinidad and tobago*. *The Economic Journal 120*(549), 1399–1429.

Kane, T. J. and D. O. Staiger (2008). Estimating teacher impacts on student achievement: An experimental evaluation. Working Paper 14607, National Bureau of Economic Research.

Katz, L. F., J. R. Kling, and J. B. Liebman (2001, May). Moving to opportunity in boston: Early results of a randomized mobility experiment. *The Quarterly Journal of Economics 116*(2), 607–654.

Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature 48*(2), 281–355.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology 79*(3), 280–295.

Mcclain, M.-C. and S. 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, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714.

Oakes, J. (2005). *Keeping Track: How Schools Structure Inequality, Second Edition* (2 edition ed.). New Haven, Conn.; London: Yale University Press.

Panel, N. M. A. (2008). Foundations for success: Report of the national mathematics advisory panel. Technical report, U.S. Department of Education, Washington, D.C. Archived: President Bush issued an executive order creating a National Mathematics Advisory Panel to advise him and Secretary of Education Margaret Spellings on the best use of scientifically based research on the teaching and learning of math.

Petrilli, M. J. (2011, October). Unrealized potential. *The New York Times*.

Pop-Eleches, C. and M. Urquiola (2013). Going to a better school: Effects and behavioral responses. *American Economic Review 103*(4), 1289–1324.

Rickles, J. H. (2013, February). Examining heterogeneity in the effect of taking algebra in eighth grade. *The Journal of Educational Research 106*(4), 251–268.

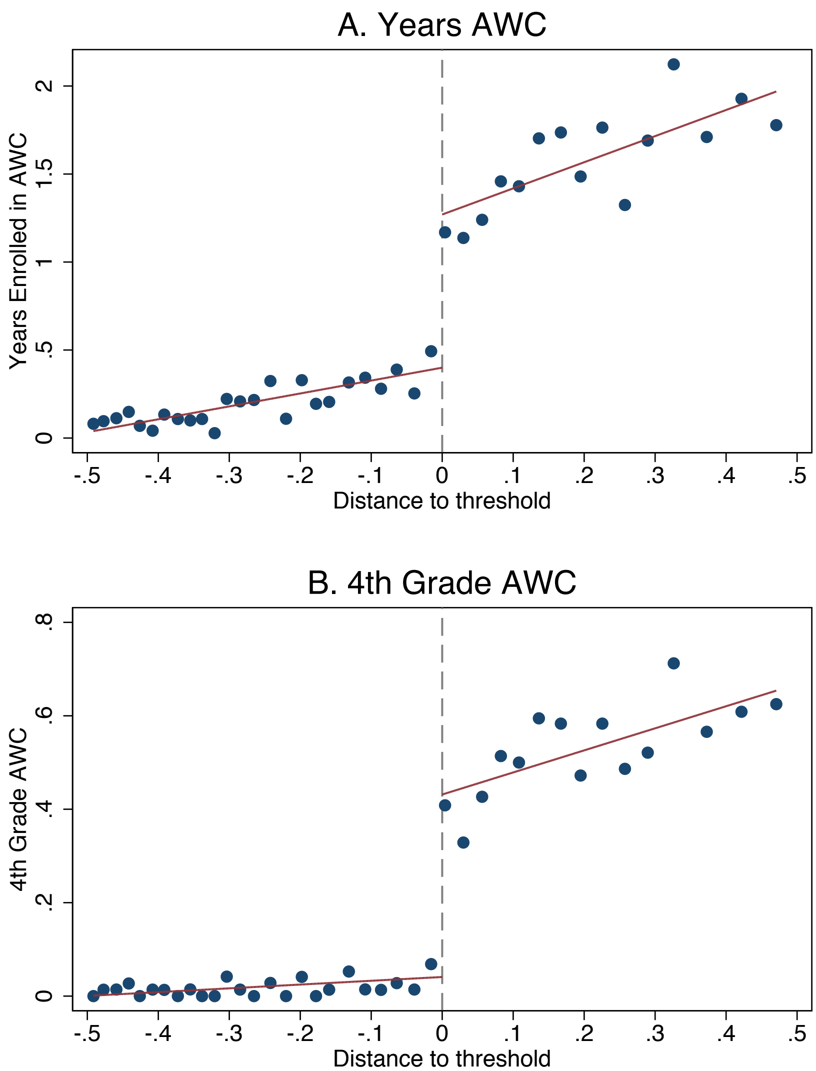
Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review 94*(2), 247–252.

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In S. M. a. L. W. Eric A. Hanushek (Ed.), *Handbook of the Economics of Education*, Volume Volume 3, pp. 249–277. Elsevier.

Stein, M. K., J. H. Kaufman, M. Sherman, and A. F. Hillen (2011). Algebra a challenge at the crossroads of policy and practice. *Review of Educational Research 81*(4), 453–492.

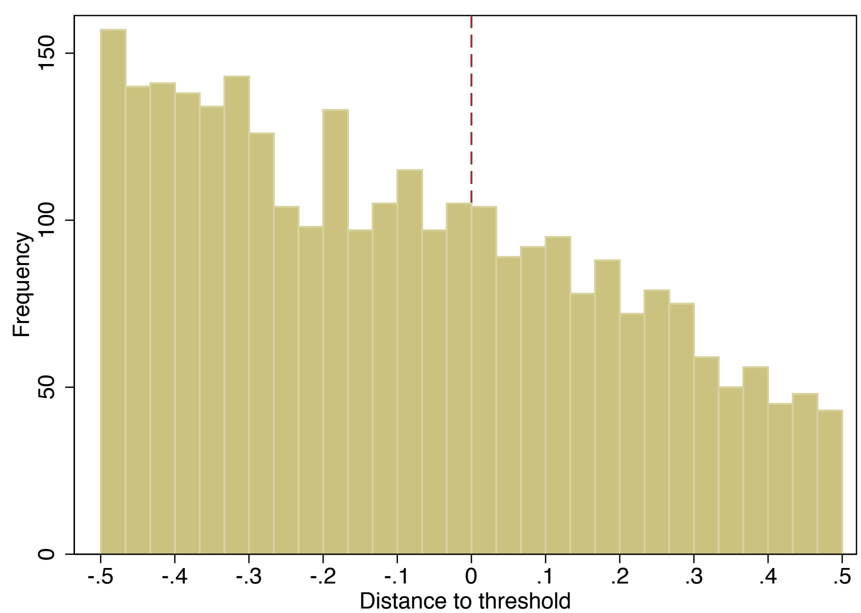
Taylor, E. (2014). Spending more of the school day in math class: Evidence from a regression discontinuity in middle school. *Journal of Public Economics 117*, 162–181.

Figure 1: AWC Enrollment by Distance to Eligibility Threshold



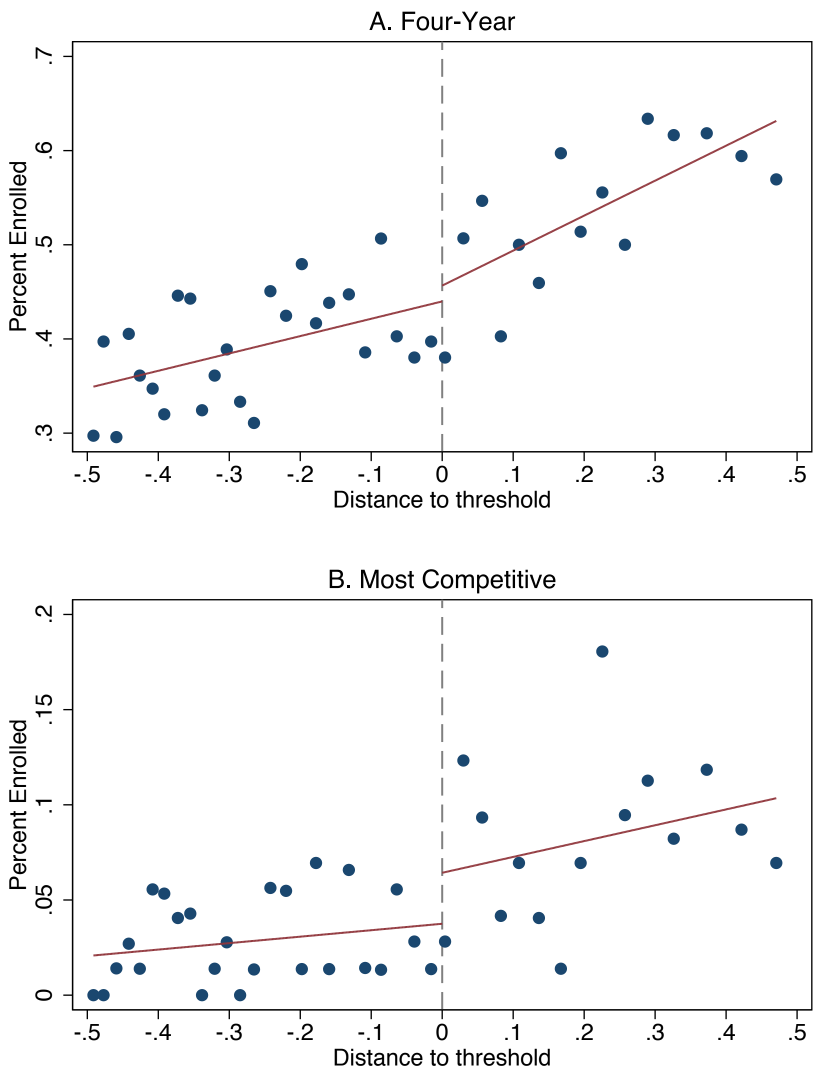
Notes: The above figure shows AWC enrollment by the running variable for the 3rd grade cohorts from 2001 to 2003 within the bandwidth of 0.5. Each dot represents the average enrollment for a bin of width 0.025. Panel A shows years of AWC enrollment, which can range between 0 and 3, and Panel B shows enrollment in 4th grade AWC.

Figure 2: Distribution of Scores near the Threshold



Notes: The above figure shows the distribution of the running variable for the third 3rd cohorts from 2001 to 2003 within the bandwidth of 0.5. The running variable is the distance of a student's combined math and reading Stanford 9 scores from a given year's AWC threshold.

Figure 3: On Time College Enrollment



Notes: The above figure shows college enrollment of students by the running variable for the 3rd grade cohorts from 2001 to 2003 within the bandwidth of 0.5. Each dot represents the average of the college enrollment rate for a bin of width 0.025.

Table 1: Descriptive Statistics

	All Students (1)	Enrolled in 4th Grade AWC (2)	RD Sample (3)	Students Above 0.5 (4)
(A) Demographics				
Female	0.481	0.517	0.513	0.527
Black	0.495	0.238	0.373	0.136
Hispanic	0.291	0.197	0.222	0.085
White	0.122	0.257	0.212	0.430
Asian	0.086	0.302	0.187	0.349
Other Race	0.006	0.006	0.006	0.000
Subsidized Lunch	0.839	0.675	0.757	0.488
English Language Learner	0.149	0.143	0.087	0.105
Special Education	0.194	0.014	0.043	0.023
3rd Grade ELA MCAS	-0.743	0.573	0.250	0.905
(B) AWC Enrollment				
4th Grade AWC	0.063	1.000	0.209	0.694
5th Grade AWC	0.063	0.923	0.207	0.686
6th Grade AWC	0.091	0.794	0.298	0.717
Years AWC	0.217	2.717	0.713	2.097
N	12,835	807	2,906	258

Mean values of each variable are shown by sample. Column (1) is the full sample of 3rd graders enrolled in BPS in the fall years from 2001-2003. Column (2) restricts that sample to students enrolled in AWC in 4th grade. Column (3) restricts the full sample to those within 0.5 of the eligibility threshold. Column (4) restricts the full sample to those more than 0.5 units away from the eligibility threshold.

Table 2: Outcome Means

	All Students (1)	Enrolled in 4th Grade AWC (2)	RD Sample (3)	Students Above 0.5 (4)
(A) 4th Grade MCAS				
ELA	-0.647	0.578	0.228	0.955
Math	-0.572	0.720	0.296	1.104
Writing Composition	-0.349	0.572	0.266	0.801
Writing Topic Development	-0.267	0.607	0.205	0.865
N	11,858	798	2,720	249
(B) 10th Grade MCAS				
ELA	-0.436	0.610	0.286	0.857
Math	-0.340	0.956	0.482	1.224
Science	-0.470	0.648	0.220	0.996
Writing Composition	-0.311	0.467	0.196	0.519
Writing Topic Development	-0.281	0.351	0.105	0.498
N	9,048	667	2,207	201
(C) High School Milestones				
Took Any AP	0.223	0.620	0.409	0.698
Took SAT	0.424	0.726	0.599	0.717
4-Year graduation	0.436	0.716	0.593	0.725
5-Year graduation	0.555	0.778	0.673	0.760
N	12,835	807	2,906	258
(D) College Enrollment within 6 mos.				
Any College	0.331	0.643	0.510	0.659
4-Year College	0.247	0.600	0.444	0.643
Most Competitive	0.021	0.105	0.048	0.178
2-Year College	0.085	0.043	0.066	0.016
N	12,835	807	2,906	258

Mean values of each outcome are shown by sample. Column (1) is the full sample of 3rd graders enrolled in BPS in the fall years from 2001-2003. Column (2) restricts that sample to students enrolled in AWC in 4th grade. Column (3) restricts the full sample to those within 0.5 of the eligibility threshold. Column (4) restricts the full sample to those more than 0.5 units away from the eligibility threshold.

Table 3: First Stage Estimates of Years of AWC Enrollment

	4th Grade (1)	5th Grade (2)	6th Grade and Above (3)
Years AWC	0.379***	0.684***	0.834***
	(0.034)	(0.068)	(0.097)
$\bar{Y}$	0.065	0.181	0.439
N	2,906	2,906	2,906

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* p<.10 ** p<.05 *** p<.01). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2003. Listed below each 2SLS coefficient is the mean of the outcome for students between 0 and 0.05 units below the eligibility threshold.

Table 4: Fuzzy Regression Discontinuity Estimates of Effects on MCAS Academic Indices and Class Rank

	Academic Index			Class Rank (Percentile)		
	Elementary School	Middle School	10th Grade	Elementary School	Middle School	10th Grade
	(1)	(2)	(3)	(4)	(5)	(6)
(A) All Students						
Reduced Form	0.025	0.016	0.060	-1.274	1.287	2.348
	(0.046)	(0.043)	(0.051)	(1.726)	(1.617)	(2.699)
2SLS	0.044	0.019	0.070	-2.294	1.505	2.676
	(0.082)	(0.050)	(0.057)	(3.098)	(1.857)	(3.056)
CCM	0.125	0.423	0.354	67.194	65.428	55.462
N	5,349	7,292	2,322	5,348	7,281	2,173
(B) Low-Income Students						
Reduced Form	-0.001	0.015	0.040	-2.519	-0.022	0.262
	(0.056)	(0.051)	(0.065)	(1.995)	(1.913)	(3.102)
2SLS	-0.001	0.018	0.050	-4.500	-0.026	0.324
	(0.100)	(0.060)	(0.079)	(3.537)	(2.291)	(3.794)
CCM	0.050	0.390	0.334	66.279	67.077	54.842
N	4,073	5,616	1,759	4,072	5,608	1,638
(C) Minority Students						
Reduced Form	0.019	0.013	0.101	-2.480	2.607	3.413
	(0.066)	(0.063)	(0.081)	(2.195)	(2.150)	(4.610)
2SLS	0.038	0.018	0.143	-4.951	3.516	4.810
	(0.130)	(0.084)	(0.109)	(4.414)	(2.969)	(6.367)
CCM	0.172	0.400	0.380	75.770	71.255	67.049
N	3,135	4,212	1,324	3,135	4,205	1,197

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2003. Listed below each 2SLS coefficient is the control complier mean. The academic index is the mean of all available MCAS subject test z-scores, standardized to be mean zero, standard deviation one. Elementary school regressions stack 4th and 5th grade outcomes, include grade fixed effects, and double cluster standard errors by 3rd grade school and student. Middle school regressions stack 6th, 7th, and 8th grade outcomes, include grade fixed effects, and double cluster standard errors by 3rd grade school and student.

Table 5: Fuzzy Regression Discontinuity Estimates of Effects on Math Course Sequence

	Prealg. by 7th (1)	Algebra 1 by 8th (2)	Algebra 2 by 9th (3)	Geometry by 10th (4)	Precalc by 11th (5)	Calculus by 12th (6)
(A) All Students						
2SLS	-0.032	0.120**	0.065	0.039	0.031	0.022
	(0.056)	(0.053)	(0.063)	(0.042)	(0.053)	(0.036)
CCM	0.385	0.597	0.527	0.675	0.598	0.171
N	3,924	4,055	3,986	3,910	3,792	3,850
(B) Low-Income Students						
2SLS	-0.015	0.102*	0.089	0.095	0.041	0.034
	(0.065)	(0.056)	(0.078)	(0.059)	(0.066)	(0.045)
CCM	0.420	0.637	0.583	0.640	0.555	0.208
N	2,852	2,961	2,970	2,946	2,881	2,909
(C) Minority Students						
2SLS	-0.014	0.122	0.085	0.053	0.044	-0.035
	(0.072)	(0.084)	(0.087)	(0.076)	(0.068)	(0.047)
CCM	0.401	0.509	0.425	0.495	0.538	0.173
N	2,437	2,493	2,458	2,360	2,256	2,297

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* p<.10 ** p<.05 *** p<.01). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools who match to student course data (2011-2013). The fall 3rd grade cohorts vary by grade level of the outcome: 7th grade, 2006-2009; 8th grade, 2005-2008; 9th grade, 2004-2007; 10th grade, 2003-2006; 11th grade, 2002-2005; 12th grade, 2001-2004.

Table 6: Fuzzy Regression Discontinuity Estimates of Effects on High School Milestones

	Took Any AP (1)	Score 3+ Any AP (2)	Took AP Calc (3)	Score 3+ AP Calc (4)	Took SAT (5)	Score MA Med.+ SAT (6)	Four-year Grad. (7)	Five-year Grad. (8)
(A) All Students								
2SLS	0.091**	-0.013	0.046*	-0.012	-0.042	0.016	0.034	-0.006
	(0.035)	(0.032)	(0.027)	(0.017)	(0.038)	(0.049)	(0.045)	(0.044)
CCM	0.482	0.271	0.065	0.038	0.724	0.378	0.680	0.746
N	2,899	2,899	2,899	2,899	2,899	2,899	2,899	2,899
(B) Low-Income Students								
2SLS	0.075	-0.026	0.068*	-0.007	-0.040	0.034	0.064	0.009
	(0.050)	(0.042)	(0.038)	(0.024)	(0.057)	(0.054)	(0.063)	(0.062)
CCM	0.497	0.288	0.098	0.052	0.701	0.356	0.630	0.703
N	2,185	2,185	2,185	2,185	2,185	2,185	2,185	2,185
(C) Minority Students								
2SLS	0.058	-0.035	0.035	-0.003	0.019	-0.014	0.128**	0.068
	(0.056)	(0.050)	(0.042)	(0.027)	(0.060)	(0.066)	(0.062)	(0.067)
CCM	0.364	0.171	0.043	0.017	0.626	0.241	0.522	0.571
N	1,718	1,718	1,718	1,718	1,718	1,718	1,718	1,718

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* p<.10 ** p<.05 *** p<.01). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2003.

Table 7: Fuzzy Regression Discontinuity Estimates of Effects on College Enrollment within 6 Months of Expected High School Graduation

	Any (1)	Four-year (2)	Four-year Private (3)	Four-year Public (4)	Most Competitive (5)	Two-year (6)
(A) All Students						
2SLS	0.057	0.019	0.042	-0.023	0.042**	0.038
	(0.046)	(0.044)	(0.040)	(0.043)	(0.020)	(0.029)
CCM	0.598	0.520	0.211	0.308	0.020	0.078
N	2,899	2,899	2,899	2,899	2,899	2,899
(B) Low-Income Students						
2SLS	0.048	0.018	0.013	0.005	0.040	0.030
	(0.050)	(0.051)	(0.051)	(0.052)	(0.024)	(0.038)
CCM	0.671	0.594	0.260	0.333	0.030	0.078
N	2,185	2,185	2,185	2,185	2,185	2,185
(C) Minority Students						
2SLS	0.097	0.060	0.130*	-0.070	0.043	0.036
	(0.075)	(0.070)	(0.070)	(0.061)	(0.029)	(0.038)
CCM	0.533	0.422	0.148	0.274	-0.002	0.111
N	1,718	1,718	1,718	1,718	1,718	1,718

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* p<.10 ** p<.05 *** p<.01). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2003. Listed below each 2SLS coefficient is the control complier mean. College quality determined by the 2009 Barron's rankings.

Table 8: Robustness Checks, 2SLS Coefficients

	Years AWC (FS) (1)	ES Ac. In. (2)	MS Ac. In. (3)	HS Ac. In. (4)	Alg1 by 8th (5)	Took Any AP (6)	Took AP Calc (7)	4-yr HS Grad (8)	Ontime Enroll 4 yr (9)	Ontime Most Comp. (10)
(A) Reference										
Baseline	0.834***	0.044	0.019	0.070	0.120**	0.091**	0.046*	0.034	0.019	0.042**
	(0.097)	(0.082)	(0.050)	(0.057)	(0.053)	(0.035)	(0.027)	(0.045)	(0.044)	(0.020)
(B) Specifications										
No controls	0.822***	0.023	-0.013	0.038	0.121**	0.080**	0.045	0.025	0.006	0.042**
	(0.093)	(0.090)	(0.048)	(0.055)	(0.056)	(0.035)	(0.029)	(0.045)	(0.044)	(0.020)
Official (2003)	0.820***	0.096	0.046	0.012	0.146**	0.116	0.066	0.137	-0.133	0.054
	(0.097)	(0.141)	(0.096)	(0.152)	(0.065)	(0.093)	(0.067)	(0.098)	(0.130)	(0.047)
No 2001	0.789***	-0.013	-0.014	0.011	0.120**	0.051	0.026	-0.001	-0.101	0.026
	(0.134)	(0.091)	(0.066)	(0.084)	(0.053)	(0.052)	(0.042)	(0.063)	(0.072)	(0.027)
Quadratic	1.007***	0.086	0.046	0.099**	0.158***	0.060*	0.027	0.024	0.008	0.037**
	(0.082)	(0.068)	(0.040)	(0.049)	(0.042)	(0.031)	(0.025)	(0.029)	(0.033)	(0.017)
CCT	0.549***	0.092	0.121	0.477**	0.262**	0.391***	0.054	0.289 **	0.074	0.086**
	(0.145)	(0.129)	(0.097)	(0.207)	(0.123)	(0.146)	(0.055)	(0.137)	(0.086)	(0.042)
BW	0.21	0.29	0.22	0.18	0.24	0.17	0.31	0.17	0.30	0.27
(C) Bandwidths										
BW = 0.75	0.891***	0.017	-0.002	0.042	0.112***	0.067**	0.014	0.030	0.017	0.033**
	(0.090)	(0.060)	(0.039)	(0.048)	(0.037)	(0.029)	(0.023)	(0.031)	(0.032)	(0.016)
BW = 0.25	0.750***	0.120	0.104	0.230***	0.200**	0.130*	0.035	0.106	0.008	0.076**
	(0.127)	(0.121)	(0.074)	(0.080)	(0.084)	(0.067)	(0.038)	(0.076)	(0.085)	(0.034)
IK bandwidth	0.822***	0.019	0.000	0.053	0.114***	0.065**	0.048*	0.032	0.019	0.038**
	(0.099)	(0.063)	(0.038)	(0.051)	(0.036)	(0.027)	(0.029)	(0.024)	(0.040)	(0.018)
BW	0.45	0.72	0.83	0.63	0.78	0.86	0.45	1.17	0.58	0.59

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). First stage estimates of years enrolled in AWC are from the non-MCAS (thus full sample) outcomes. All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation, except for the rows labeled No controls which exclude these controls. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5, except for where the bandwidth is otherwise labeled or for the rows labeled Quadratic, which include the full sample, a rectangular kernel, and a second order polynomial. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2003, except for the algebra 1 by 8th grade outcome, which includes 3rd grade cohorts from 2005-2008.

Table 9: Fuzzy Regression Discontinuity Estimates of Effects on 4th through 6th Grade Classroom Characteristics

	Black (1)	Hispanic (2)	White (3)	Asian (4)
(A) Peers				
2SLS (AWC 4th)	-0.078***	-0.096***	0.101***	0.074***
	(0.022)	(0.023)	(0.025)	(0.016)
CCM	0.291	0.343	0.141	0.173
N	9,594	9,594	9,594	9,594
	Subsidized lunch (5)	Eng. lang. learner (6)	Special education (7)	3rd grade MCAS (8)
(B) Peers continued				
2SLS (AWC 4th)	-0.135***	-0.099***	-0.066***	0.657***
	(0.027)	(0.020)	(0.012)	(0.061)
CCM	0.783	0.328	0.109	-0.174
N	9,594	9,594	9,594	9,594
	ELA VA (10)	Math VA (11)	Years Exp. (12)	Novice (13)
(C) Teachers				
2SLS (AWC 4th)	0.053	0.102	0.388	-0.059**
	(0.171)	(0.143)	(1.090)	(0.025)
CCM	0.354	0.233	10.321	0.097
N	8,133	7,971	9,356	9,594
	Academic Index (14)	Class Rank School (15)	Class Rank Classroom (16)	
(D) MCAS Comparison				
2SLS (Years AWC)	0.061	-0.437	-12.803***	
	(0.064)	(2.192)	(2.153)	
CCM	0.291	67.001	59.089	
N	9,537	9,537	9,536	

Notes: Robust clustered standard errors are in parentheses (* p<.10 ** p<.05 *** p<.01). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2007 to 2012 in the grade levels that student-teacher-class links are available. Listed below each coefficient is the control complier mean. Third grade MCAS is the average of math and ELA scores. Regressions stack 4th, 5th grade, and 6th grade outcomes, include grade fixed effects, and triple cluster standard errors by 3rd grade school, classroom, and student.

Table 10: 2SLS Estimates of Treatment Channels, Multi-Site Instrument

	Academic Index						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(A) Value-Added							
Years AWC	0.026			-0.031	0.001		-0.014
	(0.031)			(0.057)	(0.029)		(0.058)
First stage F-statistic	151.3			48.8	217.7		63.7
Peer Scores		0.099		0.161		0.016	0.045
		(0.103)		(0.193)		(0.096)	(0.185)
First stage F-statistic		191.5		142.8		61.4	236.6
Value-Added			0.240***		0.240***	0.237**	0.236**
			(0.082)		(0.087)	(0.094)	(0.096)
First stage F-statistic			59.7		40.0	58.5	33.2
(B) Novice Teacher							
Years AWC	0.026			-0.031	0.009		-0.025
	(0.031)			(0.057)	(0.034)		(0.055)
First stage F-statistic	151.3			48.8	131.3		83.0
Peer Scores		0.099		0.161		0.049	0.099
		(0.103)		(0.193)		(0.116)	(0.201)
First stage F-statistic		191.5		142.8		138.1	161.4
Novice Teacher			-0.963		-0.924	-0.838	-0.822
			(0.636)		(0.686)	(0.743)	(0.751)
First stage F-statistic			7.5		0.8	6.0	77.7

Notes: Robust clustered standard errors are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). N = 8731. All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. Value-added is the standardized average of math and ELA value-added. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2007 to 2012 in the grade levels that student-teacher-class links are available, in the classrooms that it is possible to estimate leave-year-out value-added scores. Regressions stack 4th, 5th grade, and 6th grade outcomes, include grade fixed effects, and triple cluster standard errors by 3rd grade school, classroom, and student.

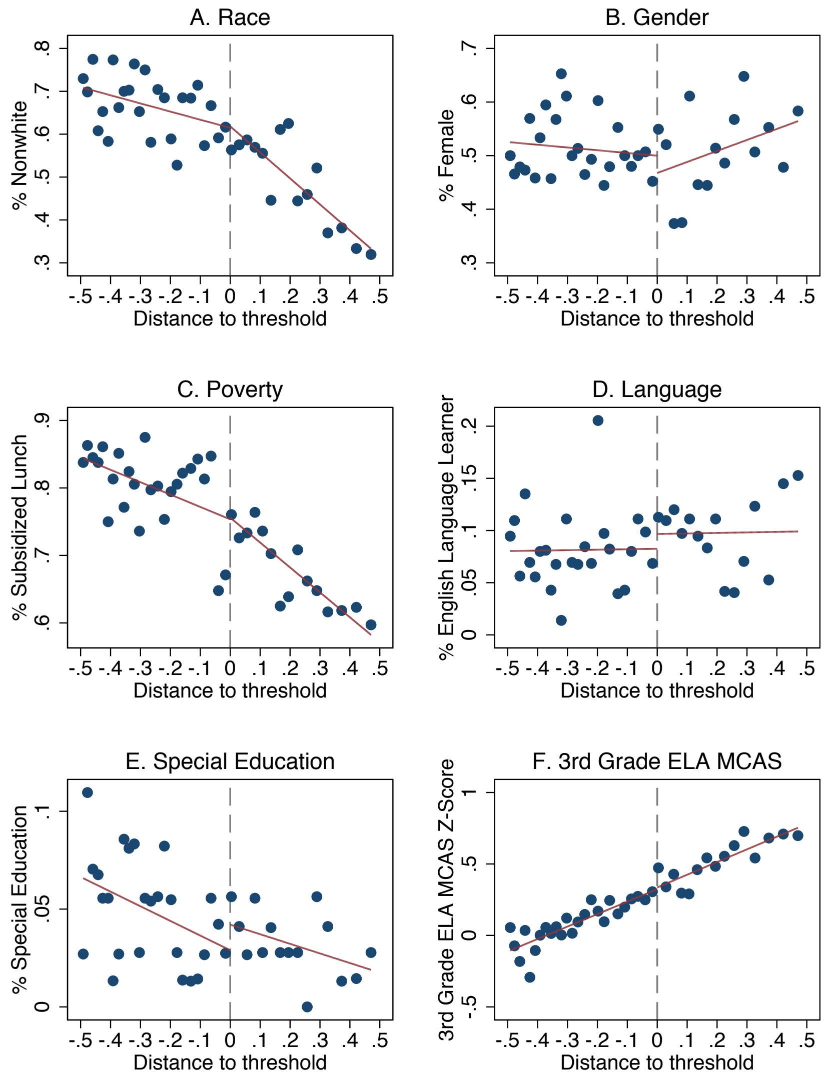
Table 11: 2SLS Estimates of Treatment Channels, College Outcomes, Multi-Site Instrument

Endogenous Variable(s)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(A) 4-Year College							
Years AWC	0.066***						0.034
	(0.020)						(0.026)
First stage F-statistic	49.7						23.5
Peer Scores		0.333***					0.060
		(0.089)					(0.126)
First stage F-statistic		95.0					55.0
Took Any Non-Calc AP			0.412***				-0.027
			(0.094)				(0.107)
First stage F-statistic			25.1				19.3
Took AP Calculus				0.304**			0.109
				(0.127)			(0.121)
First stage F-statistic				193.8			150.4
Took SAT					0.610***		0.380***
					(0.068)		(0.116)
First stage F-statistic					27.8		15.6
Grad. H.S. On-Time						0.592***	0.259*
						(0.089)	(0.133)
First stage F-statistic						120.2	84.5
(B) Most Competitive							
Years AWC	0.005						0.020
	(0.014)						(0.020)
First stage F-statistic	49.7						23.5
Peer Scores		-0.064					-0.129
		(0.073)					(0.104)
First stage F-statistic		95.0					55.0
Took Any Non-Calc AP			0.129**				0.119
			(0.056)				(0.080)
First stage F-statistic			25.1				19.3
Took AP Calculus				0.248***			0.210**
				(0.087)			(0.085)
First stage F-statistic				193.8			150.4
Took SAT					0.038		-0.007
					(0.058)		(0.083)
First stage F-statistic					27.8		15.6
Grad. H.S. On-Time						0.026	-0.043
						(0.047)	(0.072)
First stage F-statistic						120.2	84.5

Notes: Robust clustered standard errors are in parentheses (* p<.10 ** p<.05 *** p<.01). N = 2899. All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2003.

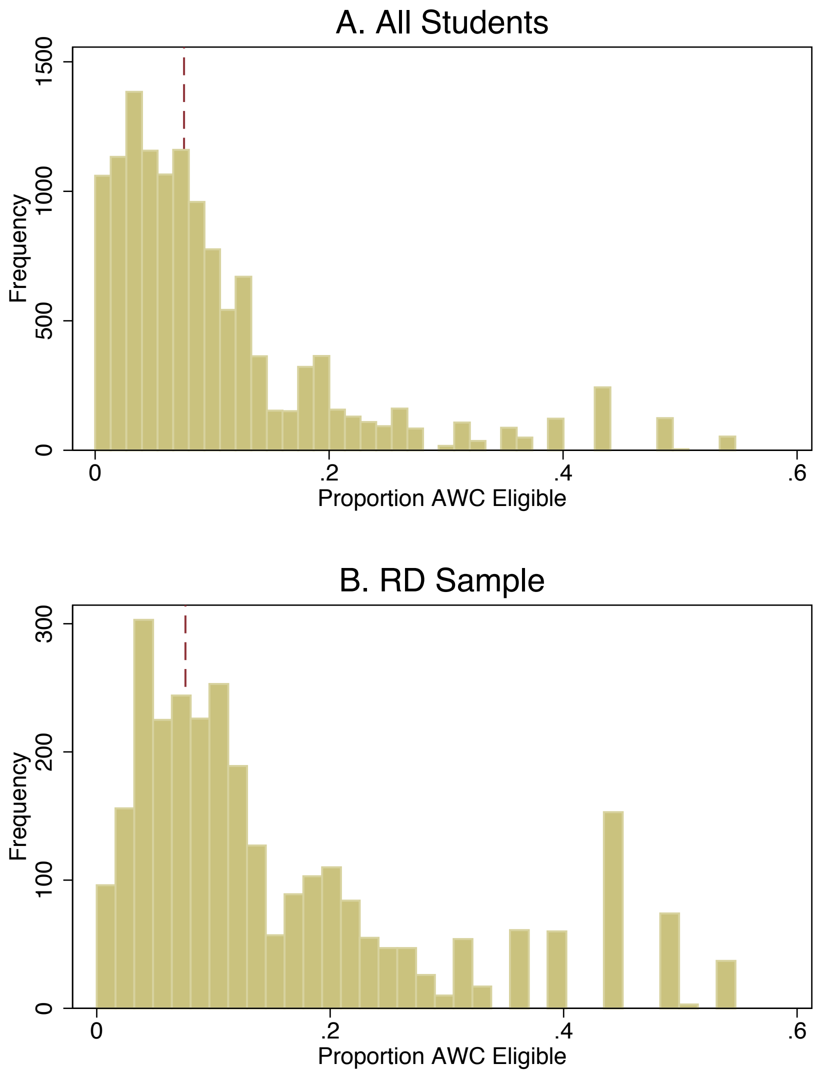
Appendix A: Additional Results

Figure A.1: Covariate Balance



Notes: The above figure shows descriptive characteristics of students by the running variable for the 3rd grade cohorts from 2001 to 2003 within the bandwidth of 0.5. Each dot represents the average of the descriptive characteristics for a bin of width 0.025.

Figure A.2: Distribution of School-Level Proportion AWC Eligible



Notes: The above figure shows the distribution of school-level AWC eligiblity rates, at the student observation level. Panel A shows this distribution for all students from the 3rd grade cohorts of 2001-2003 and Panel B limits to those within 0.5 of the AWC eligiblity threshold. The dotted line indicates the median eligbility rate, which is 0.076.

Table A.1: Covariate Balance by AWC Eligibility

	Female (1)	Black (2)	Hispanic (3)	Asian (4)	Subsidized lunch (5)	Eng. Lang. Learner (6)	Special ed. (7)	3rd grade MCAS ELA (8)
AWC Eligibility	-0.010	0.006	-0.023	-0.033	0.025	0.035	0.013	0.015
	(0.048)	(0.035)	(0.028)	(0.029)	(0.042)	(0.027)	(0.018)	(0.052)
$\bar{Y}$	0.477	0.387	0.219	0.194	0.658	0.084	0.039	0.293
N	2,906	2,906	2,906	2,906	2,906	2,906	2,906	2,850

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). Demographic controls listed at the top of each column are used as outcomes. All regressions include 3rd grade school by year fixed effects. Each coefficient on AWC eligibility is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2003. Listed below each 2SLS coefficient is the mean of the outcome for students between 0 and 0.05 units below the eligibility threshold.

Table A.2: Attrition: Fuzzy Regression Discontinuity Estimates of Effects on Leaving the Sample

	4th Grade (1)	5th Grade (2)	6th Grade (3)	7th Grade (4)	8th Grade (5)	9th Grade (6)	10th Grade (7)	11th Grade (8)	12th Grade (9)	Not Sent to NSC (10)
Reduced Form	-0.008	-0.010	-0.051	-0.013	-0.037	-0.014	-0.003	0.041	0.022	-0.002
	(0.015)	(0.022)	(0.032)	(0.039)	(0.035)	(0.028)	(0.034)	(0.036)	(0.039)	(0.028)
2SLS	-0.022	-0.015	-0.061*	-0.016	-0.045	-0.016	-0.004	0.049	0.027	-0.002
	(0.038)	(0.030)	(0.036)	(0.044)	(0.041)	(0.032)	(0.039)	(0.043)	(0.044)	(0.031)
CCM	0.029	0.022	0.117	0.174	0.167	0.197	0.216	0.211	0.195	0.057
N	2,899	2,899	2,899	2,899	2,899	2,899	2,899	2,899	2,899	2,899

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2003. Listed below each 2SLS coefficient is the control complier mean.

Table A.3: Characteristics of Students who Take-Up AWC, by AWC Eligibility

	Below Threshold (1)	Above Threshold (2)
Female	-0.001	-0.122
	(0.005)	(0.091)
Black	-0.038***	-0.016
	(0.012)	(0.168)
Hispanic	-0.026**	0.247*
	(0.013)	(0.147)
Asian	0.015	0.651***
	(0.020)	(0.170)
Other Race	-0.074***	0.371
	(0.015)	(0.433)
Subsidized Lunch	-0.014	0.009
	(0.010)	(0.151)
English Language Learner	0.047***	0.014
	(0.015)	(0.156)
Special education	-0.014***	-0.593***
	(0.004)	(0.198)
3rd Grade ELA MCAS	0.044***	0.297***
	(0.004)	(0.060)
Constant	0.124***	1.386***
	(0.018)	(0.260)
R-squared	0.044	0.065
N	11,049	1,307

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* p<.10 ** p<.05 *** p<.01). The outcome is years of AWC enrollment. All student characteristics are measured in 3rd grade. The excluded group are male, white students who do not participate in the subsidized lunch, special education or English language learner programs. All columns are restricted to 3rd graders enrolled in BPS in the fall years from 2001-2003. Column (1) restricts this sample further to those below eligibility threshold for AWC. Column (2) restricts this sample further to those above eligibility threshold for AWC.

Table A.4: Descriptive Statistics for the Regression Discontinuity Sample, by AWC Take-Up

	Below Threshold No AWC (1)	Below Threshold AWC (2)	Above Threshold AWC (3)	Above Threshold No AWC (4)
(A) Demographics				
Female	0.508	0.541	0.504	0.526
Black	0.426	0.365	0.269	0.357
Hispanic	0.253	0.206	0.188	0.172
White	0.173	0.216	0.243	0.316
Asian	0.143	0.213	0.290	0.150
Other Race	0.006	0.000	0.010	0.005
Subsidized Lunch	0.814	0.747	0.689	0.657
English Language Learner	0.076	0.108	0.107	0.079
Special Education	0.056	0.014	0.021	0.052
3rd Grade ELA MCAS	0.028	0.399	0.559	0.455
(B) AWC Enrollment				
4th Grade AWC	0.000	0.118	0.808	0.000
5th Grade AWC	0.000	0.196	0.769	0.000
6th Grade AWC	0.000	0.939	0.830	0.000
Years AWC	0.000	1.253	2.407	0.000
N	1,536	296	707	367

Mean values of each variable are shown by sample. All columns are restricted to 3rd graders enrolled in BPS in the fall years from 2001-2003 within 0.5 of the threshold. Column (1) restricts this sample further to those below eligibility threshold who do not enroll in AWC. Column (2) restricts this sample further to those below eligibility threshold who do enroll in AWC. Column (3) restricts this sample further to those above eligibility threshold who do enroll in AWC. Column (4) restricts this sample further to those above eligibility threshold who do not enroll in AWC.

Table A.5: Outcome Means for the Regression Discontinuity Sample, by AWC Take-Up

	Below Threshold No AWC	Below Threshold AWC	Above Threshold AWC	Above Threshold No AWC
(A) 4th Grade MCAS				
ELA	-0.014	0.475	0.541	0.385
Math	0.016	0.560	0.686	0.448
Writing Composition	0.117	0.371	0.517	0.274
Writing Topic Development	0.058	0.220	0.494	0.212
N	1,414	295	699	312
(B) 10th Grade MCAS				
ELA	0.062	0.487	0.571	0.412
Math	0.187	0.724	0.901	0.557
Science	-0.029	0.425	0.582	0.286
Writing Composition	0.047	0.253	0.427	0.255
Writing Topic Development	-0.032	0.206	0.306	0.138
N	1,116	249	594	248
(C) High School Milestones				
Took Any AP	0.309	0.514	0.605	0.365
Took SAT	0.530	0.709	0.734	0.537
4-Year graduation	0.532	0.676	0.713	0.553
5-Year graduation	0.622	0.777	0.777	0.605
N	1,536	296	707	367
(D) College Enrollment within 6 mos.				
Any College	0.435	0.598	0.644	0.496
4-Year College	0.363	0.530	0.595	0.420
Most Competitive	0.022	0.061	0.089	0.068
2-Year College	0.072	0.068	0.048	0.076
N	1,536	296	707	367

Mean values of each outcome are shown by sample. All columns are restricted to 3rd graders enrolled in BPS in the fall years from 2001-2003 within 0.5 of the threshold. Column (1) restricts this sample further to those below eligibility threshold who do not enroll in AWC. Column (2) restricts this sample further to those below eligibility threshold who do enroll in AWC. Column (3) restricts this sample further to those above eligibility threshold who do enroll in AWC. Column (4) restricts this sample further to those above eligibility threshold who do not enroll in AWC.

Table A.6: Fuzzy Regression Discontinuity Estimates of Effects on Enrollment

	4th Grade (1)	5th Grade (2)	6th Grade (3)	7th Grade (4)	8th Grade (5)	9th Grade (6)	10th Grade (7)	11th Grade (8)	12th Grade (9)
(A) BPS Schools									
2SLS (All)	-0.044	-0.020	0.058	-0.007	0.020	-0.009	-0.003	-0.022	0.015
	(0.049)	(0.046)	(0.041)	(0.051)	(0.051)	(0.046)	(0.047)	(0.049)	(0.049)
CCM	0.996	0.963	0.767	0.701	0.685	0.688	0.656	0.635	0.644
2SLS (Exam)	-	-	-	-0.039	-0.024	-0.026	-0.020	-0.001	-0.013
	-	-	-	(0.041)	(0.038)	(0.042)	(0.046)	(0.045)	(0.047)
CCM	-	-	-	0.366	0.351	0.424	0.441	0.421	0.420
(B) Boston Charter Schools									
2SLS	0.014	0.004	-0.022	-0.028	-0.025	-0.019	-0.007	-0.007	0.006
	(0.013)	(0.022)	(0.024)	(0.022)	(0.021)	(0.018)	(0.017)	(0.019)	(0.015)
CCM	-0.005	0.006	0.053	0.048	0.059	0.027	0.015	0.022	0.012
(C) Other MA Public Schools									
2SLS	0.052	0.031	0.025	0.051	0.049	0.044	0.013	-0.020	-0.048
	(0.046)	(0.034)	(0.032)	(0.036)	(0.040)	(0.043)	(0.041)	(0.039)	(0.038)
CCM	-0.020	0.009	0.063	0.078	0.089	0.088	0.112	0.132	0.150
N	2,899	2,899	2,899	2,899	2,899	2,899	2,899	2,899	2,899

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2003. Listed below each 2SLS coefficient is the control complier mean. Students who live in Boston as 3rd graders can enroll in non-Boston public schools in two ways. One is to reside in the district and participate in METCO, a program that allows out-of-district enrollment, the other is to move out of the district. In the regression discontinuity sample, the vast majority of students attend out of Boston schools through moving, not through METCO.

Table A.7: Fuzzy Regression Discontinuity Estimates of Effects on Exam School Application

	Apply Any Exam (1)	Apply BLS (2)	Apply BLA (3)	Apply O'Bryant (4)	Offer Any Exam (5)	Offer BLS (6)	Offer BLA (7)	Offer O'Bryant (8)	ISEE Z-Score (9)	GPA Z-Score (10)
(A) 7th Grade										
Reduced Form	0.015	0.019	0.016	0.006	0.002	0.012	-0.007	-0.003	0.078	-0.087
	(0.036)	(0.036)	(0.036)	(0.035)	(0.039)	(0.022)	(0.024)	(0.024)	(0.077)	(0.078)
2SLS	0.018	0.022	0.019	0.008	0.002	0.014	-0.008	-0.004	0.078	-0.087
	(0.040)	(0.040)	(0.040)	(0.040)	(0.045)	(0.025)	(0.028)	(0.028)	(0.067)	(0.070)
CCM	0.669	0.672	0.668	0.645	0.418	0.109	0.203	0.106	0.261	0.069
N	2,899	2,899	2,899	2,899	2,899	2,899	2,899	2,899	1,270	1,270
(B) 9th Grade										
Reduced Form	0.034	0.032	0.025	0.026	0.023	0.023*	-0.007	0.007	0.174	0.284*
	(0.045)	(0.044)	(0.045)	(0.046)	(0.030)	(0.012)	(0.015)	(0.024)	(0.173)	(0.171)
2SLS	0.041	0.038	0.030	0.031	0.027	0.028**	-0.009	0.008	0.204	0.334*
	(0.052)	(0.051)	(0.051)	(0.053)	(0.035)	(0.014)	(0.017)	(0.027)	(0.134)	(0.177)
CCM	0.273	0.276	0.234	0.203	0.130	0.033	0.045	0.052	0.402	0.339
N	2,899	2,899	2,899	2,899	2,899	2,899	2,899	2,899	489	489

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$).All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2003. Listed below each 2SLS coefficient is the control complier mean. BLS stands for Boston Latin School, BLA for Boston Latin Academy, and O'Bryant for John D. O'Bryant School of Math and Science.

Table A.8: Fuzzy Regression Discontinuity Estimates of Effects on MCAS Scores

	ELA (1)	Math (2)	Science (3)	Writing Composition (4)	Writing Topic Development (5)
(A) Elementary School					
Reduced Form	0.032	-0.000	0.024	-0.014	0.021
	(0.075)	(0.065)	(0.094)	(0.082)	(0.107)
2SLS	0.045	0.061	-0.001	0.050	0.108
	(0.108)	(0.107)	(0.098)	(0.171)	(0.180)
CCM	0.172	0.299	-0.034	0.305	0.192
N	3,610	3,622	2,601	2,712	2,712
(B) Middle School					
Reduced Form	0.083	0.010	0.058	-0.095	-0.153
	(0.074)	(0.072)	(0.097)	(0.092)	(0.113)
2SLS	0.040	0.012	-0.016	-0.037	-0.057
	(0.057)	(0.059)	(0.078)	(0.078)	(0.087)
CCM	0.306	0.546	-0.096	0.501	0.401
N	6,396	7,265	2,363	2,410	2,410
(C) 10th Grade					
Reduced Form	0.097	0.090	0.073	0.006	0.135
	(0.093)	(0.079)	(0.087)	(0.113)	(0.110)
2SLS	0.082	0.075	-0.013	0.057	0.104
	(0.060)	(0.063)	(0.068)	(0.075)	(0.087)
CCM	0.222	0.545	0.291	0.322	0.204
N	2,200	2,192	2,264	2,200	2,200

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2003. Listed below each 2SLS coefficient is the control complier mean. Elementary school regressions stack 4th and 5th grade outcomes, include grade fixed effects, and double cluster standard errors by 3rd grade school and student. Middle school regressions stack 6th, 7th, and 8th grade outcomes, include grade fixed effects, and double cluster standard errors by 3rd grade school and student.

Table A.9: Fuzzy Regression Discontinuity Estimates of Effects on Advanced Placement Test Taking and Scores

	Any AP (1)	Any English (2)	U.S. Hist or Gov't (3)	Any Science (4)	Any Calculus (5)
(A) Took AP Exam					
Reduced Form	0.076**	0.020	0.017	-0.004	0.039
	(0.033)	(0.029)	(0.029)	(0.030)	(0.024)
2SLS	0.091**	0.023	0.020	-0.004	0.046*
	(0.035)	(0.033)	(0.033)	(0.034)	(0.027)
CCM	0.482	0.278	0.172	0.150	0.065
(B) Scored above 3 on AP Exam					
Reduced Form	-0.010	-0.023	0.002	-0.014	-0.010
	(0.028)	(0.028)	(0.025)	(0.019)	(0.015)
2SLS	-0.013	-0.028	0.003	-0.017	-0.012
	(0.032)	(0.032)	(0.029)	(0.023)	(0.017)
CCM	0.271	0.143	0.081	0.054	0.038
(C) Scored above 4 on AP Exam					
Reduced Form	-0.038	-0.007	0.010	-0.013	-0.002
	(0.025)	(0.020)	(0.020)	(0.017)	(0.013)
2SLS	-0.045	-0.009	0.012	-0.016	-0.003
	(0.030)	(0.023)	(0.023)	(0.019)	(0.015)
CCM	0.145	0.052	0.018	0.001	0.048
N	2,899	2,899	2,899	2,899	2,899

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* p<.10 ** p<.05 *** p<.01). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2003. Listed below each 2SLS coefficient is the control complier mean.

Table A.10: Fuzzy Regression Discontinuity Estimates of Effects on SAT Test Taking and Scores

	Composite (2400) (1)	Verbal (800) (2)	Math (800) (3)	Writing (800) (4)
(A) Took SAT				
Reduced Form	-0.035	-	-	-
	(0.033)	-	-	-
2SLS	-0.042	-	-	-
	(0.038)	-	-	-
CCM	0.724	-	-	-
(B) Scored above MA Median				
Reduced Form	0.013	-0.015	-0.026	0.005
	(0.044)	(0.033)	(0.039)	(0.041)
2SLS	0.016	-0.018	-0.032	0.006
	(0.049)	(0.038)	(0.046)	(0.047)
CCM	0.378	0.355	0.495	0.355
N	2,899	2,899	2,899	2,899
(C) Average Score (for Takers)				
Reduced Form	14.947	4.012	4.545	6.390
	(21.165)	(7.838)	(7.537)	(9.330)
2SLS	14.587	3.915	4.436	6.236
	(18.902)	(7.027)	(6.711)	(8.387)
CCM	1549	500	543	506
N	1,722	1,722	1,722	1,722

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* p<.10 ** p<.05 *** p<.01). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2003. Listed below each 2SLS coefficient is the control complier mean.

Table A.11: Fuzzy Regression Discontinuity Estimates of Effects on MCAS Academic Indices and Class Rank, Baseline Scores Substituted for Missing Scores

	Academic Index			Class Rank (Percentile)		
	Elementary School	Middle School	10th Grade	Elementary School	Middle School	10th Grade
(A) All Students						
Reduced Form	0.017	0.002	0.047	-0.996	0.381	1.123
	(0.045)	(0.041)	(0.047)	(1.690)	(1.488)	(2.533)
2SLS	0.032	0.003	0.057	-1.869	0.456	1.343
	(0.084)	(0.048)	(0.054)	(3.163)	(1.752)	(2.998)
CCM	0.153	0.428	0.400	67.410	68.359	61.562
N	5,745	8,620	2,867	5,744	8,618	2,865
(B) Low-Income Students						
Reduced Form	-0.009	0.002	0.037	-2.099	-0.909	0.007
	(0.058)	(0.049)	(0.056)	(2.052)	(1.783)	(2.997)
2SLS	-0.016	0.003	0.047	-3.912	-1.149	0.009
	(0.107)	(0.061)	(0.069)	(3.805)	(2.277)	(3.761)
CCM	0.078	0.389	0.355	66.186	68.516	60.159
N	4,357	6,535	2,163	4,356	6,533	2,161
(C) Minority Students						
Reduced Form	0.008	-0.010	0.087	-1.842	0.481	0.522
	(0.064)	(0.056)	(0.067)	(2.167)	(1.664)	(3.505)
2SLS	0.016	-0.013	0.117	-3.827	0.642	0.696
	(0.131)	(0.074)	(0.086)	(4.510)	(2.199)	(4.623)
CCM	0.213	0.446	0.463	75.419	75.649	74.388
N	3,412	5,123	1,698	3,411	5,121	1,696

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* p<.10 ** p<.05 *** p<.01). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2003. Listed below each 2SLS coefficient is the control complier mean. The academic index is the mean of all available MCAS subject test z-scores, standardized to be mean zero, standard deviation one. Elementary school regressions stack 4th and 5th grade outcomes, include grade fixed effects, and double cluster standard errors by 3rd grade school and student. Middle school regressions stack 6th, 7th, and 8th grade outcomes, include grade fixed effects, and double cluster standard errors by 3rd grade school and student.

Table A.12: Fuzzy Regression Discontinuity Estimates of Effects on College Enrollment within 6 Months of Expected High School Graduation, 2001 Cohort (All Students Sent to NSC)

	Any (1)	Four-year (2)	Four-year Private (3)	Four-year Public (4)	Most Competitive (5)	Two-year (6)
(A) All Students						
2SLS	0.147**	0.181***	0.135*	0.046	0.063**	-0.034
	(0.073)	(0.070)	(0.072)	(0.055)	(0.029)	(0.037)
CCM	0.617	0.476	0.175	0.301	-0.014	0.141
N	1,013	1,013	1,013	1,013	1,013	1,013
(B) Low-Income Students						
2SLS	0.150	0.214**	0.161*	0.053	0.090**	-0.064
	(0.092)	(0.089)	(0.089)	(0.072)	(0.040)	(0.050)
CCM	0.766	0.587	0.207	0.380	0.012	0.179
N	748	748	748	748	748	748
(C) Minority Students						
2SLS	0.216	0.247*	0.193	0.053	0.095	-0.030
	(0.139)	(0.134)	(0.121)	(0.092)	(0.067)	(0.077)
CCM	0.610	0.363	0.062	0.301	-0.089	0.247
N	600	600	600	600	600	600

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* p<.10 ** p<.05 *** p<.01). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001. Listed below each 2SLS coefficient is the control complier mean. College quality determined by the 2009 Barron's rankings.

Table A.13: Heterogeneity by 3rd Grade School Characteristics, 2SLS Coefficients

	ES Ac. In. (1)	MS Ac. In. (2)	HS Ac. In. (3)	Alg1 by 8th (4)	Took Any AP (5)	Took AP Calc (6)	4-yr HS Grad (7)	Ontime Enroll 4 yr (8)	Ontime Most Comp. (9)
(A) Has AWC									
AWC * has AWC	0.145	0.133***	0.116**	0.112	0.095**	-0.023	0.070	0.062	0.071***
	(0.093)	(0.048)	(0.058)	(0.071)	(0.048)	(0.035)	(0.056)	(0.049)	(0.022)
AWC * no AWC	-0.005	-0.036	0.053	0.124*	0.089*	0.084**	0.015	-0.001	0.027
	(0.119)	(0.069)	(0.082)	(0.074)	(0.048)	(0.037)	(0.061)	(0.061)	(0.026)
p (AWC = no AWC)	0.330	0.044	0.544	0.909	0.925	0.037	0.513	0.424	0.195
N (AWC)	1485	2061	682	1069	799	799	799	799	799
N (No AWC)	3864	5233	1647	2990	2107	2107	2107	2107	2107
(B) Peer Quality									
AWC * high peers	0.097	0.057	-0.035	0.041	0.160***	0.023	0.072	-0.011	0.044
	(0.120)	(0.080)	(0.108)	(0.078)	(0.059)	(0.045)	(0.071)	(0.090)	(0.030)
AWC * low peers	0.026	0.004	0.167**	0.178**	0.046	0.063	0.009	0.046	0.042
	(0.105)	(0.056)	(0.069)	(0.070)	(0.046)	(0.039)	(0.056)	(0.052)	(0.028)
p (High = Low)	0.647	0.552	0.115	0.195	0.157	0.544	0.470	0.602	0.977
N (High)	2469	3346	1069	1732	1322	1322	1322	1322	1322
N (Low)	2880	3948	1260	2327	1584	1584	1584	1584	1584
(C) Eligibility Rate									
AWC * high elig.	0.153*	0.061	0.041	0.112*	0.096**	0.050*	0.044	0.008	0.024
	(0.093)	(0.059)	(0.063)	(0.058)	(0.038)	(0.027)	(0.054)	(0.055)	(0.025)
AWC * low elig.	-0.114	-0.059	0.187	0.216	0.100	0.034	0.028	0.076	0.096*
	(0.157)	(0.091)	(0.122)	(0.192)	(0.081)	(0.051)	(0.078)	(0.076)	(0.049)
p (High = Low)	0.127	0.258	0.284	0.618	0.967	0.771	0.860	0.466	0.204
N (High)	3609	4893	1565	3271	1958	1958	1958	1958	1958
N (Low)	1740	2401	764	788	948	948	948	948	948

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* p<.10 ** p<.05 *** p<.01). For details on the specification, see notes of Table A.8 (for elementary MCAS outcomes), Table 6 (for high school outcomes), and Table 6 (for college outcomes). Each panel fully interacts that baseline specification with indicators for the given categories. Below each panel is the p-value from a test of the equality of the two listed coefficients.

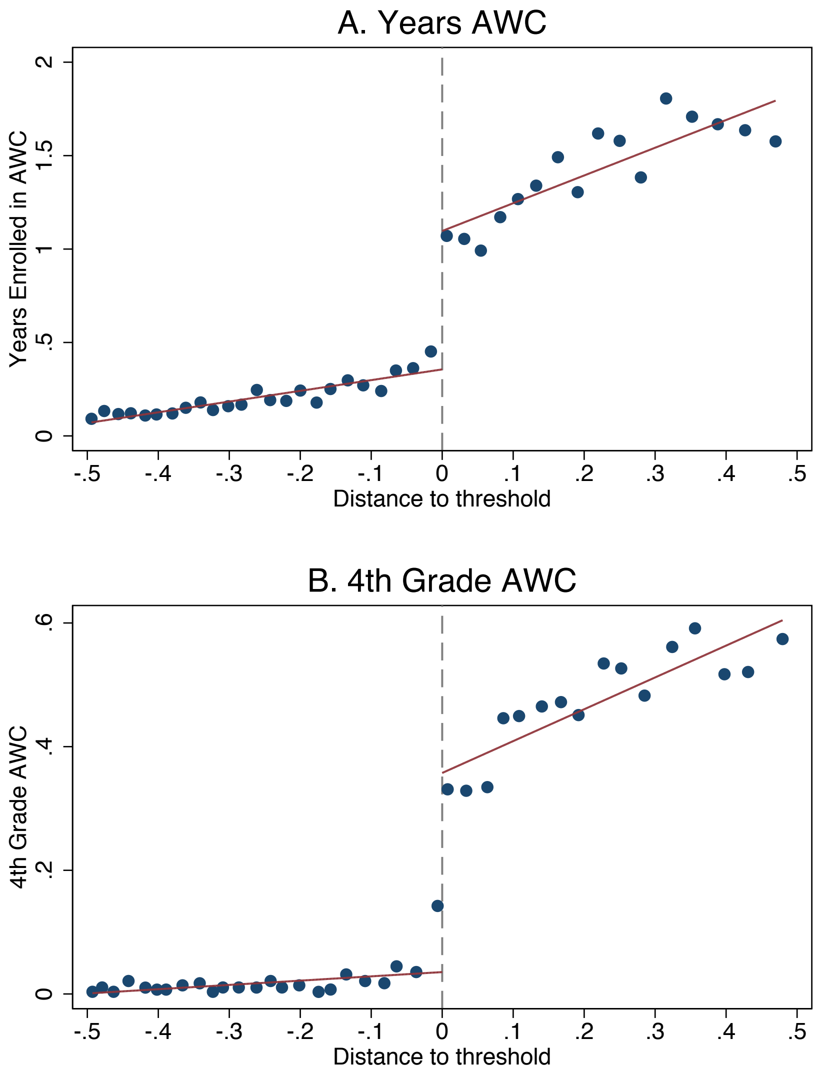
Table A.14: Descriptive Statistics for the Regression Discontinuity Sample, by 3rd GradeSchool Type

	3rd Grade School					
	Has AWC (1)	No AWC (2)	High Peers (3)	Low Peers (4)	High Eligibility (5)	Low Eligibility (6)
(A) Demographics						
Female	0.471	0.529	0.512	0.513	0.509	0.520
Black	0.310	0.396	0.279	0.451	0.306	0.509
Hispanic	0.155	0.247	0.163	0.271	0.190	0.288
White	0.175	0.226	0.287	0.150	0.263	0.109
Asian	0.355	0.123	0.266	0.121	0.236	0.085
Other Race	0.004	0.007	0.005	0.008	0.005	0.008
Subsidized Lunch	0.792	0.743	0.694	0.809	0.716	0.841
English Language Learner	0.168	0.057	0.105	0.073	0.092	0.077
Special Education	0.024	0.050	0.045	0.040	0.040	0.047
3rd Grade ELA MCAS	0.162	0.283	0.325	0.187	0.276	0.197
(B) AWC Enrollment						
4th Grade AWC	0.334	0.161	0.239	0.183	0.248	0.127
5th Grade AWC	0.328	0.161	0.238	0.181	0.243	0.134
6th Grade AWC	0.378	0.267	0.332	0.269	0.326	0.239
Years AWC	1.040	0.589	0.809	0.633	0.817	0.500
N	799	2,107	1,322	1,584	1,958	948

Mean values of each variable are shown by sample. All columns are restricted to 3rd graders enrolled in BPS in the fall years from 2001-2003 within 0.5 of the threshold. Column (1) restricts this sample further to those whose 3rd grade school hosts an AWC program. Column (2) restricts this sample further to those whose 3rd grade school does not host an AWC program. Column (3) restricts this sample further to those whose 3rd grade school has average 3rd grade test scores greater than or equal to -0.5σ. Column (4) restricts this sample further to those whose 3rd grade school has average 3rd grade test scores below -0.5σ. Column (5) restricts this sample further to those whose 3rd grade school has an AWC eligiblity rate greater than or equal to 7.6 percent. Column (6) restricts this sample further to those whose 3rd grade school has an AWC eligiblity rate below to 7.6 percent.

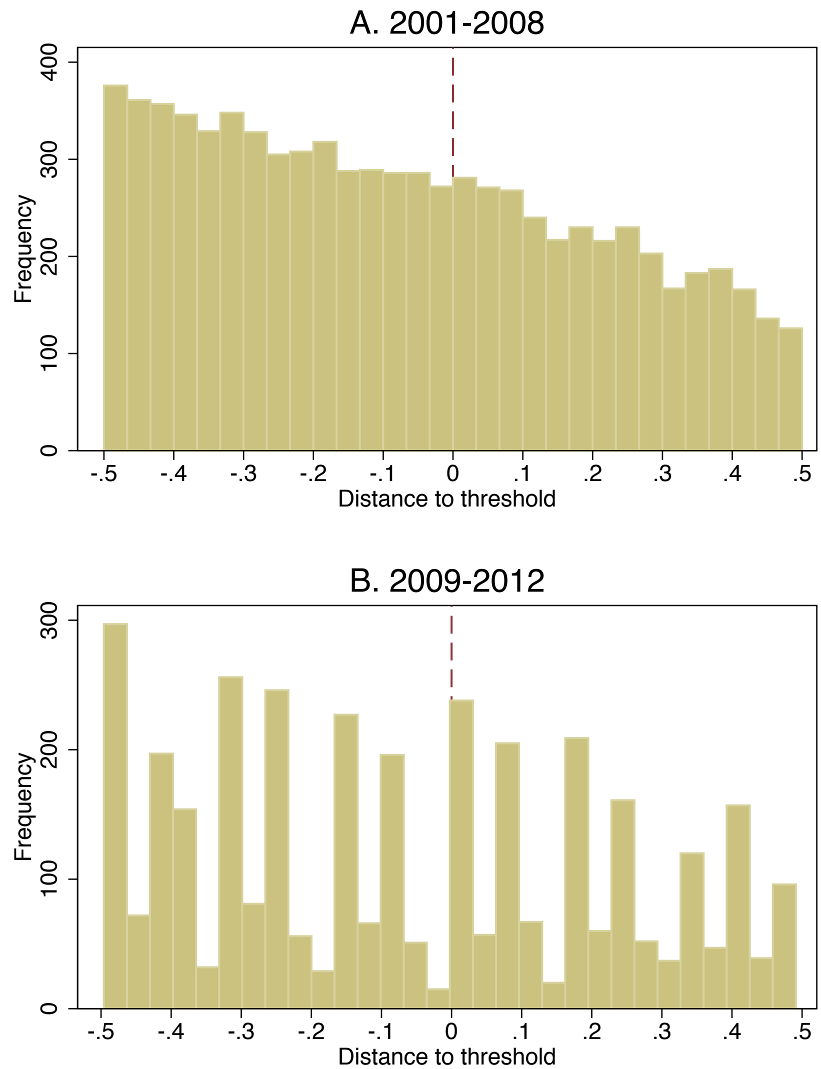
Appendix B: Main Results Using All Available 3rd Grade Cohorts

Figure B.1: AWC Enrollment by Distance to Eligibility Threshold, All Cohorts



Notes: The above figure shows AWC enrollment by the running variable for the third grade cohorts within the bandwidth of 0.5. Each dot represents the average enrollment for a bin of width 0.025. Panel A shows years of AWC enrollment, which can range between 0 and 3, and is limited to 3rd grade cohorts from fall 2001-2010 to allow students to reach the maximum potential of years of AWC enrollment, and Panel B shows enrollment in 4th grade AWC for the 3rd grade cohorts from 2001-2012.

Figure B.2: Distribution of Scores near the Threshold, All Cohorts



Notes: The above figure shows the distribution of the running variable within the bandwidth of 0.5. Panel A shows the 3rd grade cohorts from 2001 to 2008 who were tested with the Stanford 9 exam. Panel B shows the 3rd grade cohorts from 2009-2012, who were tested with the TerraNova. The TerraNova has fewer points than the Stanford 9, which explains the pronounced sawtooth pattern observed in Panel B. The running variable is the distance of a student's combined math and reading Stanford 9/TerraNova scores from a given year's AWC threshold.

Table B.1: Descriptive Statistics, All Cohorts

	All Students (1)	Enrolled in 4th Grade AWC (2)	RD Sample (3)
(A) Demographics			
Female	0.483	0.524	0.507
Black	0.412	0.223	0.320
Hispanic	0.351	0.219	0.287
White	0.126	0.256	0.209
Asian	0.080	0.269	0.151
Other Race	0.031	0.033	0.033
Subsidized Lunch	0.821	0.629	0.736
English Language Learner	0.260	0.176	0.182
Special Education	0.199	0.023	0.058
3rd Grade ELA MCAS	-0.693	0.555	0.214
(B) AWC Enrollment			
4th Grade AWC	0.075	1.000	0.201
5th Grade AWC	0.070	0.842	0.190
6th Grade AWC	0.085	0.630	0.233
Years AWC	0.230	2.472	0.625
N	46,221	3,469	11,458

Mean values of each variable are shown by sample. Column (1) is the full sample of 3rd graders enrolled in BPS in the fall years from 2001 to 2012. Column (2) restricts that sample to students enrolled in AWC in 4th grade. Column (3) restricts the full sample to those within 0.5 of the eligibility threshold.

Table B.2: Outcome Means, All Cohorts

	All Students (1)	Enrolled in 4th Grade AWC (2)	RD Sample (3)
(A) 4th Grade MCAS			
ELA	-0.603	0.599	0.223
Math	-0.480	0.674	0.306
Writing Composition	-0.357	0.482	0.164
Writing Topic Development	-0.311	0.480	0.116
N	43,256	3,388	10,883
(B) 10th Grade MCAS			
ELA	-0.425	0.591	0.254
Math	-0.307	0.907	0.455
Science	-0.423	0.642	0.218
Writing Composition	-0.303	0.396	0.168
Writing Topic Development	-0.291	0.316	0.083
N	16,867	1,337	4,447
(C) High School Milestones			
Took Any AP	0.231	0.616	0.419
Took SAT	0.424	0.726	0.599
4-Year graduation	0.447	0.721	0.605
5-Year graduation	0.555	0.778	0.673
N	12,835	807	2,906
(D) College Enrollment within 6 mos.			
Any College	0.331	0.643	0.510
4-Year College	0.247	0.600	0.444
Most Competitive	0.021	0.105	0.048
2-Year College	0.085	0.043	0.066
N	12,835	807	2,906

Mean values of each outcome are shown by sample. Column (1) is the full sample of 3rd graders enrolled in BPS in the fall years from 2001 to 2012. Column (2) restricts that sample to students enrolled in AWC in 4th grade. Column (3) restricts the full sample to those within 0.5 of the eligibility threshold.

Table B.3: Covariate Balance by AWC Eligibility, All Cohorts

	Female (1)	Black (2)	Hispanic (3)	Asian (4)	Subsidized lunch (5)	Eng. Lang. Learner (6)	Special ed. (7)	3rd grade MCAS ELA (8)
AWC Eligibility	-0.000	0.015	-0.037**	-0.000	0.008	0.001	0.001	-0.019
	(0.022)	(0.020)	(0.017)	(0.012)	(0.018)	(0.014)	(0.012)	(0.025)
$\bar{Y}$	0.504	0.341	0.289	0.157	0.713	0.159	0.050	0.248
N	11,458	11,458	11,458	11,458	11,458	11,458	11,458	11,293

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* p<.10 ** p<.05 *** p<.01). Demographic controls listed at the top of each column are used as outcomes. All regressions include 3rd grade school by year fixed effects. Each coefficient on AWC eligibility is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2012. Listed below each 2SLS coefficient is the mean of the outcome for students between 0 and 0.05 units below the eligibility threshold.

Table B.4: Attrition: Fuzzy Regression Discontinuity Estimates of Effects on Leaving the Sample, All Cohorts

	4th Grade (1)	5th Grade (2)	6th Grade (3)	7th Grade (4)	8th Grade (5)	9th Grade (6)	10th Grade (7)	11th Grade (8)	12th Grade (9)	Not Sent to NSC (10)
Reduced Form	-0.004	-0.011	-0.023	-0.002	-0.013	-0.029	-0.006	-0.008	0.031	-0.002
	(0.009)	(0.011)	(0.015)	(0.019)	(0.019)	(0.020)	(0.026)	(0.031)	(0.032)	(0.028)
2SLS	-0.014	-0.019	-0.032	-0.002	-0.020	-0.045	-0.008	-0.010	0.040	-0.002
	(0.026)	(0.018)	(0.020)	(0.027)	(0.028)	(0.029)	(0.033)	(0.037)	(0.039)	(0.031)
CCM	0.028	0.038	0.105	0.130	0.157	0.196	0.189	0.214	0.197	0.057
N	11,438	10,507	9,574	8,729	7,905	6,885	5,821	4,805	3,850	2,899

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2012. Listed below each 2SLS coefficient is the control complier mean.

Table B.5: First Stage Estimates of Years of AWC Enrollment, All Cohorts

	4th Grade (1)	5th Grade (2)	6th Grade (3)	7th Grade (4)	8th Grade (5)	9th Grade (6)	10th Grade (7)	11th Grade (8)	12th Grade (9)
Years AWC	0.314***	0.593***	0.710***	0.683***	0.657***	0.652***	0.744***	0.787***	0.770***
	(0.025)	(0.051)	(0.064)	(0.061)	(0.060)	(0.060)	(0.067)	(0.077)	(0.080)
$\bar{Y}$	0.058	0.161	0.400	0.422	0.422	0.433	0.479	0.444	0.441
4th Grade AWC	0.314***	0.319***	0.321***	0.311***	0.300***	0.301***	0.333***	0.356***	0.352***
	(0.025)	(0.025)	(0.025)	(0.025)	(0.024)	(0.024)	(0.027)	(0.029)	(0.030)
$\bar{Y}$	0.058	0.059	0.059	0.060	0.060	0.067	0.079	0.048	0.049
N	11,458	10,525	9,589	8,743	7,918	6,898	5,832	4,815	3,859

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* p<.10 ** p<.05 *** p<.01). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2012. Listed below each 2SLS coefficient is the mean of the outcome for students within 0.05σ below the eligibility threshold.

Table B.6: Fuzzy Regression Discontinuity Estimates of Effects on MCAS Academic Indices and Class Rank, All Cohorts

	Academic Index			Class Rank (Percentile)		
	Elementary School	Middle School	10th Grade	Elementary School	Middle School	10th Grade
	(1)	(2)	(3)	(4)	(5)	(6)
(A) All Students						
Reduced Form	0.016	0.006	0.063	-1.553	0.157	2.834
	(0.028)	(0.024)	(0.038)	(1.014)	(0.910)	(1.730)
2SLS	0.034	0.008	0.079*	-3.361	0.225	3.496
	(0.061)	(0.034)	(0.047)	(2.124)	(1.296)	(2.128)
CCM	0.238	0.375	0.362	68.808	66.724	58.081
N	20,638	22,731	4,685	20,633	22,709	4,405
(B) Low-Income Students						
Reduced Form	0.000	0.003	0.033	-2.284*	-0.512	0.633
	(0.035)	(0.033)	(0.047)	(1.190)	(1.086)	(1.755)
2SLS	0.001	0.005	0.044	-4.715**	-0.718	0.841
	(0.072)	(0.046)	(0.061)	(2.355)	(1.508)	(2.294)
CCM	0.199	0.348	0.342	68.354	67.412	58.255
N	15,242	17,116	3,552	15,238	17,100	3,315
(C) Minority Students						
Reduced Form	0.018	0.025	0.094	-1.863	0.060	0.550
	(0.035)	(0.033)	(0.060)	(1.158)	(1.298)	(2.869)
2SLS	0.038	0.037	0.141	-4.060*	0.088	0.825
	(0.076)	(0.048)	(0.089)	(2.464)	(1.889)	(4.280)
CCM	0.184	0.283	0.218	71.227	67.242	63.182
N	12,422	13,617	2,723	12,419	13,605	2,502

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* p<.10 ** p<.05 *** p<.01). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The academic index is the mean of all available MCAS subject test z-scores, standardized to be mean zero, standard deviation one. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2012. Listed below each 2SLS coefficient is the control complier mean. Elementary school regressions stack 4th and 5th grade outcomes, include grade fixed effects, and double cluster standard errors by 3rd grade school and student. Middle school regressions stack 6th, 7th, and 8th grade outcomes, include grade fixed effects, and double cluster standard errors by 3rd grade school and student.

Table B.7: Fuzzy Regression Discontinuity Estimates of Effects on High School Milestones, All Cohorts

	Took Any AP (1)	Score 3+ Any AP (2)	Took AP Calc (3)	Score 3+ AP Calc (4)	Took SAT (5)	Score MA Med.+ SAT (6)	Four-year Grad. (7)	Five-year Grad. (8)
(A) All Students								
2SLS	0.070*	-0.005	0.037*	-0.010	-0.042	0.016	0.021	-0.006
	(0.038)	(0.034)	(0.022)	(0.015)	(0.038)	(0.049)	(0.042)	(0.044)
CCM	0.477	0.251	0.054	0.029	0.724	0.378	0.693	0.746
N	3,850	3,850	3,850	3,850	2,899	2,899	3,850	2,899
(B) Low-Income Students								
2SLS	0.065	-0.017	0.052*	-0.008	-0.040	0.034	0.040	0.009
	(0.048)	(0.043)	(0.031)	(0.021)	(0.057)	(0.054)	(0.062)	(0.062)
CCM	0.507	0.280	0.079	0.042	0.701	0.356	0.671	0.703
N	2,909	2,909	2,909	2,909	2,185	2,185	2,909	2,185
(C) Minority Students								
2SLS	0.053	-0.010	0.020	-0.000	0.019	-0.014	0.076	0.068
	(0.052)	(0.041)	(0.033)	(0.021)	(0.060)	(0.066)	(0.055)	(0.067)
CCM	0.408	0.197	0.045	0.012	0.626	0.241	0.578	0.571
N	2,297	2,297	2,297	2,297	1,718	1,718	2,297	1,718

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2004 for AP and 4-year high school graduation outcomes and 2001 to 2003 for SAT and 5-year high school graduation outcomes.

Table B.8: Robustness Checks, 2SLS Coefficients

	Years AWC (FS) (1)	ES Ac. In. (2)	MS Ac. In. (3)	HS Ac. In. (4)	Alg1 by 8th (5)	Took Any AP (6)	Took AP Calc (7)	4-yr HS Grad (8)	Ontime Enroll 4 yr (9)	Ontime Most Comp. (10)
(A) Reference										
Baseline	0.710***	0.034	0.008	0.079*	0.120**	0.070*	0.037*	0.021	0.019	0.042**
	(0.064)	(0.061)	(0.034)	(0.047)	(0.053)	(0.038)	(0.022)	(0.042)	(0.044)	(0.020)
(B) Specifications										
No controls	0.707***	0.031	-0.007	0.061	0.121**	0.061	0.034	0.014	0.006	0.042**
	(0.063)	(0.063)	(0.036)	(0.049)	(0.056)	(0.041)	(0.023)	(0.042)	(0.044)	(0.020)
Official (2003+)	0.684***	0.037	0.026	0.078	0.146**	0.052	0.046	0.052	-0.133	0.054
	(0.061)	(0.071)	(0.044)	(0.073)	(0.065)	(0.088)	(0.041)	(0.076)	(0.130)	(0.047)
No 2001	0.685***	0.018	-0.001	0.058	0.120**	0.033	0.018	-0.013	-0.101	0.026
	(0.069)	(0.063)	(0.038)	(0.057)	(0.053)	(0.053)	(0.031)	(0.057)	(0.072)	(0.027)
Quadratic	0.847***	-0.007	-0.009	0.102***	0.158***	0.042	0.003	0.017	0.008	0.037**
	(0.060)	(0.041)	(0.024)	(0.036)	(0.042)	(0.032)	(0.021)	(0.029)	(0.033)	(0.017)
CCT	0.386***	0.048	0.078	0.110	0.262**	0.313***	0.046	0.200*	0.074	0.086**
	(0.061)	(0.072)	(0.051)	(0.096)	(0.123)	(0.117)	(0.054)	(0.107)	(0.086)	(0.042)
BW	0.20	0.36	0.28	0.36	0.24	0.17	0.26	0.19	0.30	0.27
(C) Bandwidths										
BW = 0.75	0.790***	0.044	0.005	0.055	0.112***	0.048	0.007	0.022	0.017	0.033**
	(0.066)	(0.041)	(0.025)	(0.038)	(0.037)	(0.030)	(0.019)	(0.031)	(0.032)	(0.016)
BW = 0.25	0.653***	0.005	0.050	0.160**	0.200**	0.106	0.036	0.063	0.008	0.076**
	(0.072)	(0.093)	(0.056)	(0.074)	(0.084)	(0.068)	(0.033)	(0.071)	(0.085)	(0.034)
IK bandwidth	0.657***	0.032	0.002	0.060	0.114***	0.047*	0.013	0.027	0.019	0.038**
	(0.070)	(0.063)	(0.027)	(0.041)	(0.036)	(0.029)	(0.020)	(0.023)	(0.040)	(0.018)
BW	0.27	0.49	0.67	0.63	0.78	0.85	0.67	1.39	0.58	0.59

Notes: Robust standard errors clustered by 3rd grade school are in parentheses (* p<.10 ** p<.05 *** p<.01). First stage estimates of years enrolled in AWC are from the non-MCAS (thus full sample) outcomes. All regressions include 3rd grade school by year fixed effects and controls for demographic characteristics and baseline program participation, except for the rows labeled No controls which exclude these controls. Each coefficient is generated by local linear regression with a triangular kernel of bandwidth 0.5, except for where the bandwidth is otherwise labeled or for the rows labeled Quadratic, which include the full sample, a rectangular kernel, and a second order polynomial. The sample is restricted to 3rd graders enrolled in Boston Public Schools in the fall of 2001 to 2012.