Essays on the Economics of Education

Essays on the Economics of Education

A dissertation presented

by

Sarah Rose Cohodes

to

The Department of Public Policy

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Public Policy

Harvard University

Cambridge, Massachusetts

May 2015

Dissertation Advisor:
Professor Christopher Avery

Author:
Sarah Rose Cohodes

Essays on the Economics of Education

Abstract

This dissertation includes three papers in the field of economics of education.

The first paper provides estimates of the long-run impacts of tracking high-achieving students using data from a Boston Public Schools (BPS) program, Advanced Work Class (AWC). AWC is an accelerated curriculum in 4th through 6th grades with dedicated classrooms. Using a fuzzy regression discontinuity approach based on the AWC entrance exam, I find that AWC has little impact on test scores. However, it improves longer-term academic outcomes including Algebra 1 enrollment by 8th grade, AP exam taking, and college enrollment. The college enrollment effect is particularly large for elite institutions. Testing potential channels for program effects provides suggestive evidence that teacher effectiveness and math acceleration account for AWC effects, with little evidence that peer effects contribute to gains.

The second paper uses item-level information from standardized tests to investigate whether large test score gains attributed to Boston charter schools can be explained by score inflation. To do so, I estimate the impact of charter school attendance on subscales of the test scores and examine them for evidence of score inflation. If charter schools are teaching to the test to a greater extent than their counterparts, one would expect to see higher scores on commonly tested standards, higher stakes subjects, and frequently tested topics. However, despite incentives to reallocate effort toward highly-tested content, and to coach to item type, I find no evidence of this type of test preparation. Boston charter middle schools perform consistently across all standardized test subscales.

The third paper analyzes a Massachusetts merit aid program that gives high-scoring students tuition waivers at in-state public colleges with lower graduation rates than available alternative colleges. A regression discontinuity design comparing students just above and below the eligibility threshold finds that students are remarkably willing to forgo college quality and that scholarship use actually lowered college completion rates. These results suggest that college quality affects college completion rates. The theoretical prediction that in-kind subsidies of public institutions can reduce consumption of the subsidized good is shown to be empirically important.

Contents

List of Tables

List of Figures

Acknowledgments

First and foremost, I thank my dissertation committee: Christopher Avery, Joshua Angrist, and Lawrence Katz have been exceptional mentors. I have benefited tremendously from their guidance and support. Their dedication to asking important questions, scientific rigor, and careful attention to detail exemplify the type of scholarship I aspire to. I was lucky to work with or be taught by each of my committee members before I began my doctoral studies, and thus I can say that they all inspired me both to start my PhD, and to finish it.

I thank Joshua Goodman, my coauthor for the third paper, for his enthusiasm and empiricism. Thanks also go to Josh for his advice and friendship throughout my doctoral studies, and for helpful conversations that influenced the first two papers of this dissertation as well. Many others have provided valuable comments and feedback on the essays in this dissertation, including Natalie Bau, David Deming, Susan Dynarski, David Figlio, Ed Glaeser, Peter Hull, Christopher Jencks, Tom Kane, Daniel Koretz, Kevin Lang, Jeff Liebman, Bridget Long, Richard Murnane, Lindsay Page, Parag Pathak, Laszlo Sandor, Marty West, John Willett, Christopher Walters, as well as seminar and conference participants at Harvard, MIT, Boston University, NBER, the Federal Reserve Banks of Boston and New York, APPAM, AEFP, and SOLE. Special thanks to Tom Dee, Jane Hannaway, Jon Fullerton, and Tom Kane, all of whom helped set me on the path that resulted in this dissertation.

For financial support, I am grateful to the Harvard Kennedy School, the Harvard University Multidisciplinary Program on Inequality and Social Policy, the Rappaport Institute for Greater Boston, and the Taubman Center for State and Local Government. The School Effectiveness and Inequality Institute at MIT and the Center for Education Policy Research at Harvard University provided institutional support and facilitated my access to data. Nicole Tateosian at HKS was always ready with the answer to a question, a laugh, or a chocolate.

I thank Carrie Conaway at the Massachusetts Department of Elementary and Secondary Education for providing access to the Massachusetts student data for all three of the papers in the dissertation. These projects would have been impossible without her support and dedication to research. Key individuals enabled additional data access for each of the three papers: for their time and assistance, I thank Kamal Chavda at the Boston Public Schools for the first essay, the

Boston area charter schools for the second essay, and Robert Lee at the Massachusetts Department of Elementary and Secondary Education for the third essay.

Thank you to my mother, Dale, and my siblings, David and Nina, for encouragement and unconditional love. Many years after his death, my father Don's intellect, humor, and devotion to his family continue to inspire me. I am honored to follow him by being the second person in our family to receive a doctoral degree from Harvard.

And finally, I thank my partner, Andy: you made this possible; you make anything possible.

In memory of Don Cohodes

Introduction

This dissertation consists of three papers in the field of economics of education. All three papers explore education policies in the state of Massachusetts and use student level records from Massachusetts. In these essays, I exploit naturally occurring quasi-random policy variation due to rules and regulations to generate estimates of causal effects of three education programs. Another shared component of the papers is the use of outcomes that go beyond, or deeper into, standardized test scores, as well as the investigation of policies that have the potential to narrow achievement gaps in the United States.

In the first paper, I provide the first estimates of the long-run impacts of tracking high-achieving students using data from the Boston Public Schools (BPS) program for high-achieving students, Advanced Work Class (AWC). Previous work on tracking high-achieving elementary and middle students in the US has shown little impact on short-run test scores; in this case, the long time horizon of AWC allows for an examination of other important outcomes. 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 to test test several potential channels for program effects suggests that teacher effectiveness and math acceleration account for AWC effects, with little evidence that peer effects contribute to gains.

The second paper uses item-level information from standardized tests to investigate whether large test score gains attributed to Boston charter schools can be explained by score inflation. Recent work has shown Boston charter schools raise standardized test scores more than their traditional school counterparts. Critics of charter schools argue that charter schools create those achievement gains by focusing exclusively on test preparation, at the expense of deeper learning. In this paper, I test that critique by estimating the impact of charter school attendance on subscales of the MCAS (Massachusetts Comprehensive Assessment System) and examining them for evidence of score inflation. If charter schools are teaching to the test to a greater extent than their counterparts, one would expect to see higher scores on commonly tested standards, higher stakes subjects, and frequently tested topics. However, despite incentives to reallocate effort away from less-frequently tested content to highly-tested content, and to coach to item type, I find no evidence of this type of test preparation. Boston charter middle schools perform consistently across all standardized test subscales.

In the third paper, coauthored with Joshua Goodman, we analyze a Massachusetts merit aid program that gives high-scoring students tuition waivers at in-state public colleges with lower graduation rates than available alternative colleges. A regression discontinuity design comparing students just above and below the eligibility threshold finds that students are remarkably willing to forgo college quality and that scholarship use actually lowered college completion rates. Specifically, scholarship eligibility induced 6.9 percent of students at the threshold to enroll in Adams colleges, but reduced the probability of earning a four-year college degree within six years by 2.5 percentage points. Half of the students induced to switch colleges would have enrolled in more competitive alternatives in the absence of the scholarship. These results suggest that college quality affects college completion rates. The theoretical prediction that in-kind subsidies of public institutions can reduce consumption of the subsidized good is shown to be empirically important.

Chapter 1

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

1.1 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

[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.

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 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]

[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.

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 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

(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.

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 1.3, I describe the data and sample and in Section 1.4 my empirical strategy. I report results in Section 1.5 and discuss potential threats to validity in Section 1.6. Section 1.7 includes a discussion of potential channels for the AWC effect and Section 1.8 concludes.

1.2 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

[4]The allocation of AWC seats was part the school desegregation plan in Boston, and those 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.

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]

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.

[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 reading, the combined national percentile is likely a little higher, but still well below the highest national achievement 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 1.1 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 offer in 3rd grade.

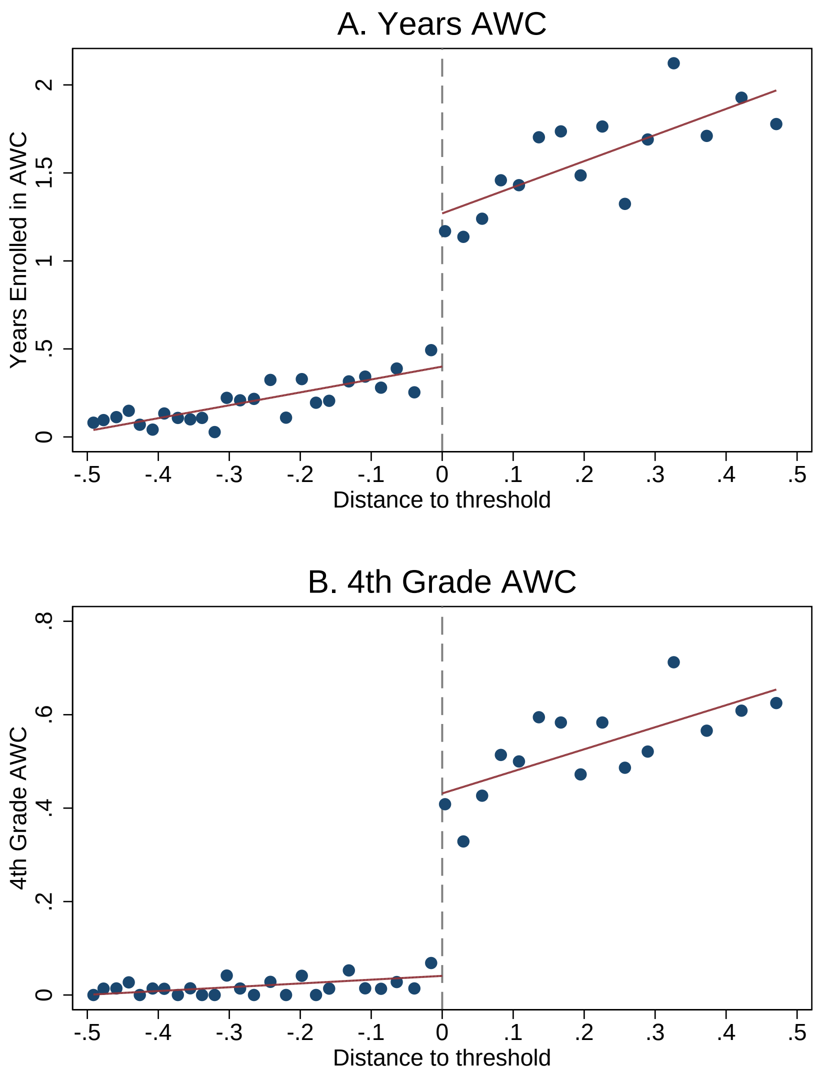
1.3 Data and Descriptive Statistics

1.3.1 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]

[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.

Figure 1.1: *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.

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 1.1 shows a discontinuous jump in years of AWC enrollment of about 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

[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.

[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.

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 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, pre-algebra; 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

[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.

[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.

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 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.

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

- *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]

- *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

[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.

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 A.2.

1.3.2 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 1.1, 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 1.1 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.

Table 1.1: *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

Notes: 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.

In terms of outcomes, I show in Table 1.2 that AWC outpace their peers in BPS. For MCAS

[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.

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 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.

1.4 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

[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.

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

Table 1.2: *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

Notes: 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.

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.

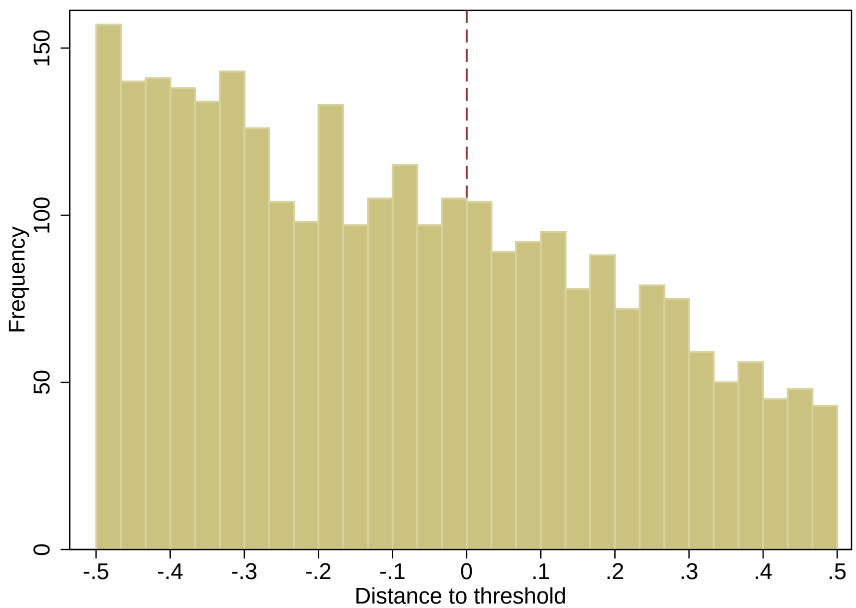
The key assumption of regression discontinuity designs is that it is impossible to manipulate scores in order to qualify for the program (McCrary, 2008). This assumption holds in the case of AWC. Since the threshold changes yearly and students do not know the algorithm that translates questions answered 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 1.2, 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 A.4 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 1.6.2, 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

[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.

Figure 1.2: *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.

described in Section 1.3.1. I create a measure of distance to the threshold, Gap, by subtracting the threshold from the combined score.[21] Figure 1.1 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 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:

$$YearsAWC_{ist+k} = \alpha_0 + \alpha_1 Above_{ist} + \alpha_2 Gap_{ist} + \alpha_3 Gap_{ist} \times Above_{ist} + \lambda' X_i + \delta_{st} + \epsilon_{ist} \tag{1.1}$$

$$Y_{ist+k} = \beta_0 + \beta_1 Years\hat{AWC}_{ist+k} + \beta_2 Gap_{ist} + \beta_3 Gap_{ist} \times Above_{ist} + \theta' X_i + \mu_{st} + \eta_{ist} \tag{1.2}$$

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

[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.

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

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 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 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

[23] I examine some of these channels in Section 1.7.

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 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:

$$\begin{aligned} Y_{ist+k} * (1 - YearsAWC_{ist+k}) &= \gamma_0 + \gamma_1(1 - Years\hat{A}WC_{ist+k}) \\ &+ \gamma_2 Gap_{ist} + \gamma_3 Gap_{ist} \times Above_{ist} + \phi' X_i + \nu_{st} + \xi_{ist} \end{aligned} \tag{1.3}$$

where $1 - YearsAWC_{ist+k}$ is instrumented by AWC eligibility as in Equation 1.2 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.

1.5 Results

1.5.1 First stage and the effect on enrollment

First stage estimates of the years of AWC enrollment are in Table 1.3. 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

[24]It is possible to estimate a treatment complier mean in a similar manner. In this case, the "TCM" is the mean 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 1.3, 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.

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

Table 1.3: *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 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.

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 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.

1.5.2 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 1.4. 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

[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 1.6.3.

[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.

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 1.3). 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.

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 A.20). 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 1.4. 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]

[28] I can do this procedure by classroom only for the more recent years of data, as shown later in Table 1.9.

Table 1.4: *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 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 school and student.

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.

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 1.5 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 pre-algebra, 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 1.5 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 National Mathematics Advisory 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

Table 1.5: *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 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.

in algebra by 8th grade is likely to be beneficial. As can be seen in Column (2) of Table 1.5, 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 1.6, 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 1.5, 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

[29]For additional subject-specific AP results, see Appendix Table A.9.

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 1.6, control complier students take the SAT at the rate of 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.

1.5.3 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 1.7, 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

[30]Colleges also accept the ACT, but most students in Massachusetts take the SAT.

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

Table 1.6: *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 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.

private universities.

Table 1.7: *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 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.

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

[32] "Most competitive" institutions include Tufts University and Boston College, the two most commonly attended

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 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 1.3. 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.

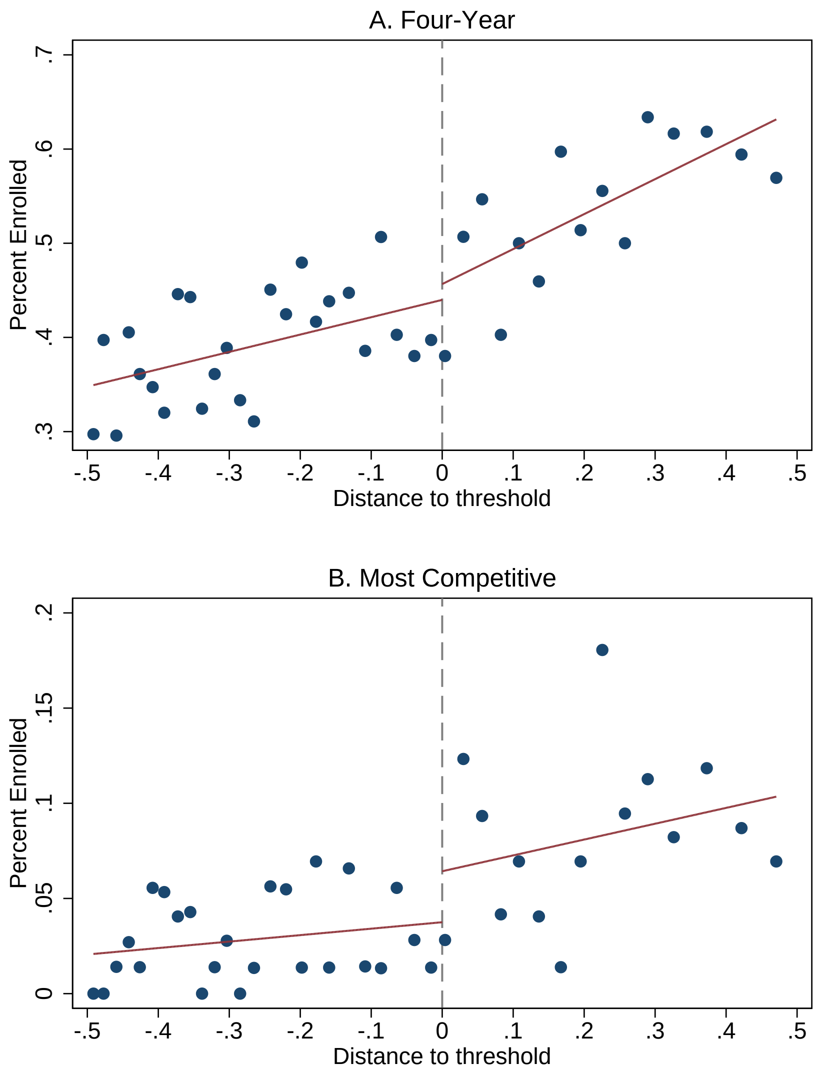
Enrollment effects are quite large for minority students. Black and Hispanic students are 10 percentage points more likely to enroll in college per year of AWC attendance, and the majority of this gain is from enrollment at four-year institutions. The switch to the private sector for four-year colleges 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.

---

highly selective institutions in my sample. It also includes the Ivy League schools and elite liberal arts colleges.

[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.

Figure 1.3: *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.

1.6 Threats to Validity

1.6.1 Robustness

The results are robust to a number of specification checks. In Table 1.8, 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 1.3.1, 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 that outcome. Here, the results are substantively the same, with an even

[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.

Table 1.8: *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 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.

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.

1.6.2 Attrition

As discussed in Section 1.4, 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 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

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

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 1.4. 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.

1.6.3 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

[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.

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.

1.7 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 1.9 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 1.1, 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 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

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

Table 1.9: *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 school, classroom, and student.

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 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

[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.

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 1.10. 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 1.9. 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

Table 1.10: *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 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 school, classroom, and student.

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 1.11. 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.

Table 1.11: *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.

1.8 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.

Chapter 2

Teaching to the Student: Charter School Effectiveness in Spite of Perverse Incentives

2.1 Introduction

Charter middle schools in Boston have obtained impressive test score results and strong reputations, resulting in hundreds of children on waitlists, hoping for a chance to enter one of these schools. According *The Boston Globe* (2011), two Boston middle school charters are in the top ten middle schools in Massachusetts, as ranked by proficiency on the 8th grade state exam. Causal research based on charter school lotteries confirmed the impressive test score results by showing that charter school students that won the lottery and attend outperform those who did not win the lottery and did not attend (Abdulkadiroğlu et al., 2009, 2011). These results are particularly important since they control for selection bias, countering the frequent criticism that charter schools "cream" certain kinds of students.

However, the mechanisms behind this large impact are unclear. Case studies and non-causal quantitative research suggest that long school days and years, small student-teacher ratios, coherent mission and curriculum, and other school characteristics may contribute to charter school success. On the other hand, another potential cause of the charter school effect is score inflation caused by test preparation activities. Score inflation is defined as “increases in scores that do not signal a commensurate increase in proficiency in the domain of interest” that the test is designed to assess (Koretz, 2008, p. 34). Two potential causes of score inflation are strategic coaching of predictable characteristics of tests and reallocation of teaching effort to highly tested topics. If charter schools are engaging in these types of activities, their strong results may be due to score inflation, rather than an actual increase in students’ comprehension. Currently, there is no quantitative evidence for or against the existence of score inflation at charter schools, but there is anecdotal evidence that charter schools are very test-aware.

The accountability system that charter schools face, which has additional accountability measures on top of NCLB, incentivizes teachers to reallocate to highly tested content and to coach certain types of items in order to raise overall score, but not necessarily increase students’ human capital. Using fine-grained data from Massachusetts, I investigate the Boston charter middle school effect more deeply to see if charter students are more successful than their counterparts in other Boston schools on all aspects of the Massachusetts Comprehensive Assessment System (MCAS) and if any of the gains can be explained by score inflation. If charter school students outpace their peers on all elements of the test—rarely tested standards and as well as common standards and topics; on science, as well as math and English/language arts (ELA); and on all types of questions (multiple choice, short answer, and open response)—then I will have no evidence of charter schools using test preparation to a greater extent that other schools in Boston.

This is the first study of charter schools that uses item-level information to disaggregate the test score effects in order to determine if charter schools are using test preparation to fuel their test score results. I present results for rarely tested content, including rarely tested curriculum standards, science, and less emphasized topics to investigate reallocation, and results by item type to investigate coaching.

Although accountability pressure from the state rating system and public competition around test score results might induce teachers to utilize test preparation, I find no evidence of this. Charter school students have large gains on almost all components of MCAS exams, leading me to suggest that their success is not due to differential test preparation, in spite of perverse incentives that might encourage it. The results are robust to adjustments made for attrition and sample matching. Additionally, charter schools do not focus on children on the “bubble” of proficiency—instead gains are magnified for the least academically prepared.

The organization of the paper is as follows. In Section 2.2, I provide the background and context by describing the charter school impact research, reviewing the relevant details of prior work in Boston, and discussing score inflation. In Section 2.3, I provide a theoretical framework. Section 2.4 describes the outcome measures, data and sample. In Section 2.5, I present my identification strategy and in Section 2.6 my results. Section 2.7 addresses threats to validity. Section 2.8 concludes.

2.2 Background and Context

2.2.1 Charter School Impacts

Lottery-based studies of charter schools have generally found positive results of charter schools on academic achievement. These studies compare students who are offered a seat at a charter school through a lottery with those that are not offered a seat, meaning that the only difference between the two groups is the random offer of charter school attendance. However, most of these lottery-based studies are small and city-specific. They are also limited to schools that are oversubscribed, which restricts their generalizability. Additionally, lottery-based results may overestimate the underlying citywide results if higher demand occurs at higher quality schools. Hoxby, Muraka, and Kang's (2009) investigation of New York City charter schools found gains for charter school students in grades 4 through 8. Dobbie and Fryer (2011) focus on one charter school in the Harlem Children's Zone in New York City and found dramatic results, with the causal effect of charter school attendance on math achievement of around a standard deviation over the course of three years in middle school. Interestingly, a recent national lottery-based evaluation of 36 charter schools found no significant effects overall, but significant gains for attendance at urban charter schools (Gleason et al., 2010).

In Boston, the causal effect of charter school attendance on middle school math scores is 0.4 standard deviations on the MCAS, and the effect on middle school ELA scores is 0.2 standard deviations on the MCAS for each year of charter school attendance. The results for high schools are similar, though slightly smaller, with about a 0.2 standard deviation gain in both ELA and math (Abdulkadiroğlu et al., 2009, 2011). The middle schools that participate in the Boston research, updated with additional years and newly opened schools, form the sample for this study.

When examining charter school impacts across Massachusetts, the Boston effect was muted (Angrist et al., 2011, 2013), but when the impacts were disaggregated by urbanicity, urban charters performed at similar levels to the Boston schools.

Results from broad comparisons between charter schools and traditional public schools are more mixed (Center for Research on Education Outcomes, 2009; Zimmer et al., 2009). The advantage of these studies is that they include students from both highly demanded and less demanded schools. However, they cannot adjust for the omitted variable bias inherent in comparing attendees at charters with those who may have never applied to a charter. A recent report finds that matching estimators can sometimes replicate lottery-based charter effects, but finds that regression and fixed effects approaches are less successful at replication, perhaps another reason for the divergence in the literature (Forstan et al., 2012). Results from both lottery-based studies and other comparisons are limited in scope to the general impact of charter school attendance on test outcomes, not the details on these outcomes or the mechanisms behind the effects.

2.2.2 Beyond Charter School Test Impacts

While the Boston results show large impacts for highly-demanded charters, the authors cannot use the test score impacts to investigate the specific mechanisms that lead to the strong results. Quantitative research on charter schools is just beginning to investigate the mechanisms behind test score impacts. To date, charter schools have almost all been treated as a "black-box" where schools produce educational achievement by undetermined mechanisms. Hoxby, Muraka, and Kang's (2009) investigation of NYC charters attempts to peek into the black-box by associating some characteristics of charter schools with their success. They find that charter schools that have a longer school year/day, more minutes of instruction in core subjects, a "small rewards/small punishment discipline" system, a performance pay structure, and/or mission statements that emphasize academic success tend to have greater test score success than charter schools without those policies (Hoxby et al., 2009, V-5). These associations should not be interpreted causally, since while they use lottery-based estimates, the connection to characteristics is descriptive. Abdulkadiroglu et al. (2011) observe that Boston charter schools have much smaller student/teacher ratios, younger teachers, and fewer in-subject licensed teachers, but again, these are descriptive, not causal, associations. Dobbie and Fryer (2013) find that positive charter school results are associated with "frequent teacher feedback, the use of data to guide instruction, high-dosage tutoring, increased instructional time, and high expectations." Angrist et. al (2013) suggests that the positive impacts for urban Massachusetts charters are partially due to demographics and partially due to adherence to a "No Excuses" philosophy. Recent case studies of five high performing charter schools in Massachusetts, including three schools in this study, found that those successful charter schools were characterized by a strong mission and a school culture dedicated to that mission; structures "that support student learning;" a focus on getting the "right" personnel; involved parents; and "classroom procedures that maximize[d] time on task and tightly link[ed] content to the Massachusetts curriculum framework" (Merseth et al., 2009, p. 228). The factors described above may be the determinants of charters success on test scores.

2.2.3 Score Inflation

Another factor that could influence charter schools' MCAS success is test preparation. If test preparation is about "working more effectively, teaching more, [and] working harder" (Koretz, 2008) then charter school test score gains might be due to an increase in these beneficial activities. But other, less benign, kinds of test preparation might be a factor in charters' MCAS success. If test preparation focuses on trivial knowledge of the test or reallocates resources to tested subjects, it could lead to score inflation. Why would potential score inflation in MCAS scores matter? If we think that MCAS outcomes are a measure of future success, not just an academic signpost during school, then test preparation and score inflation impede the inferences that we can draw from MCAS scores. To illustrate, when there is score inflation, a high math MCAS score would give a false impression of future success in math since the high score reflects test preparation rather than increased understanding of the content matter. Thus, if charter school effects are due to test preparation, the inference that they prepare students well for future math course would be false.

Score inflation can be caused by four types of test preparation: "reallocation, alignment, coaching, [and] cheating" (Koretz, 2008, p. 251). Cheating clearly undermines the purpose of testing and leads to score inflation by increasing test scores with no parallel increase in learning (Jacob and Levitt, 2003). Reallocation, alignment, and coaching are more ambiguous. Reallocation and alignment involve focusing resources and teaching on tested (or highly tested) topics and subjects, and cause score inflation when they draw efforts away from other parts of the curriculum that actually contribute to the underlying domain that the test is attempting to measure. Coaching occurs when teaching focuses on trivial aspects of the test, taking away time from meaningful content or focusing understanding of a topic in a specific format or organization. This causes score inflation by giving the impression that students comprehend the underlying domain of the test when actually they have become proficient in test taking methods or problems presented in a specific format.

Reallocation is likely widespread: with the implementation of the No Child Left Behind Act (NCLB), school districts across the nation are spending more time on tested subjects and less time on other subjects (McMurrer, 2007; Nichols and Berliner, 2007). Effects on test scores can be seen through gains on highly tested content but smaller or no gains on other content. Jacob (2005) finds that the implementation of high-stakes testing in Chicago led to gains on math items that are "easy to teach" or more common on the assessment, but no gains on other parts of the test, implying that reallocation to highly tested subjects caused the math gains. In Boston high school charters, Merseth (2010) sees impressive results on the MCAS but "less impressive results" on college entrance exams, and she suggests that teaching at the schools may focus on material in line with the state exam but not the "higher-order cognitive tasks" tested on the SAT.[1]

As mentioned above, coaching involves teaching students about test-specific aspects of the assessment, rather than content. Some familiarity with test forms is important, but techniques that teach methods of guessing or standard responses to open response questions can inflate scores. Hamilton (2003) describes case studies and nationwide studies where teachers only distribute problems that parallel the formats on the test and change their instruction to "mirror the format" of state exams. Koretz (2008) describes methods like the process of elimination on multiple choice

[1]Merseth (2010) reports 100% participation rates for taking the SAT at the three Boston charters for which she reports results (Academy of the Pacific Rim, Boston Collegiate Charter School, and MATCH). And while she reports the SAT results as "less impressive" than MCAS results, all three schools exceed the average Boston Public Schools (BPS) SAT score, even though only around 65% of Boston students take the SAT. The different compositions of who takes the test may account for the lack of a wider test score gap between the charters and BPS.

exams that, if taught, would increase students' test taking skills but not the knowledge that tests are trying to assess.

Despite their successes, Boston charters are not immune to the accountability pressures that might induce test preparation and result in score inflation. While widely perceived as successful schools because of their MCAS scores, NCLB's Adequate Yearly Progress rankings identify most Boston area charters as needing improvement. In 2011, the only Boston charter middle schools *not* identified as in "improvement" or "corrective action" status under NCLB's standards for subgroups were Edward Brooke and Excel (Massachusetts Department of Elementary and Secondary Education, 2011a). Boston charter schools, like many other schools in the nation, have the threat of NCLB sanctions as an incentive to do well on standardized exams. They are also under pressure to maintain high MCAS rankings that are widely trumpeted. Finally, charter schools must be renewed every five years in Massachusetts. While renewals are not solely based on test scores, academic achievement is part of the renewal process. These triple pressures might encourage test preparation which would cause score inflation. In Section 2.3, I describe in more detail how accountability systems can distort behavior to induce score inflation.

There is also evidence that the Boston charter schools are very test conscious. Merseth et al.'s (2009) in depth study of five charters, three of which are included in this study, indicates that teachers and administrators are very test aware. Merseth et al. report that curriculum is carefully prepared to match with the Massachusetts Curriculum Frameworks. Teachers use publicly available MCAS items from prior years and they use assessments similar to the MCAS, and teachers constantly track their students' progress on content that is tested. However, these test aware behaviors need not lead to score inflation if the test preparation activities involve teaching more or better, rather than reallocating time to tested subjects or coaching on trivial details.

2.2.4 Implications

Boston middle school charters produce large gains for their students on the MCAS. However, the mechanisms behind Boston charter middle schools' success on the MCAS are unclear. They may be due to structural reasons, like longer school days and years, or low student-teacher ratios. They may be due to curriculum and planning efforts. Or they may be due to differential test preparation that results in score inflation. The purpose of this paper is to attempt to discover more details on this apparent success. I do so by disaggregating the MCAS scores so as to separate MCAS outcomes that are susceptible to test preparation from those that are not.

By determining if charters do not perform consistently across all measures of the test, I can look for evidence of test preparation. For instance, a particularly large effect on the multiple choice outcome, but little or no effect on the open response or short answer outcomes might indicate coaching to item type. Similarly, a particularly large effect on standards that are tested most frequently, but little or no effect on standards tested rarely might indicate reallocation within mathematics to highly tested topics. For an additional check for this type of reallocation, I also exploit the fact that science is less emphasized in the accountability system and investigate whether science gains are similar in size to math and ELA gains.

2.3 Theoretical Framework

Several articles (Jacob and Levitt, 2003; Muralidharan and Sundararaman, 2011; Barlevy and Neal, 2012) have framed score inflation as a principal-agent problem. Accountability systems are put into place by state education agencies and the federal government to improve student achievement, but individual actors in the education system have an incentive to change their behavior so as to increase measured student achievement, and not necessarily students' underlying knowledge. Jacob and Levitt (2003) argue that accountability incentivizes cheating, and find overt cheating in 4 to 5 percent of Chicago classrooms. On the other hand, Muralidharan and Sundararaman (2011), argue that while a teacher incentive pay system in India might induce perverse responses, there is no evidence of such responses. Barlevy and Neal's (2012) theoretical incentive scheme also induces socially optimal responses.

Accountability systems may be formal, such as those prescribed by the No Child Left Behind Act and state educational agencies. In Massachusetts, charters face an additional accountability system with 5 year reviews from their authorizing agency – the state. In its reviews, the state requires charters to have "academic program success," "organizational viability", and "faithful to the terms of the charter." Student performance is accounted for by the academic program requirements, which, prior to 2013, included MCAS proficiency or growth towards proficiency and AYP. The factors are also accounted for in the charter faithfulness requirement, as many charter school missions include an explicit focus on academic success. While there are many other aspects of the reauthorization process, student academic achievement is quite important. Charter schools have similar pressures under the authorization process, the state accountability system, and NCLB, since they all rely on MCAS and proficiency levels or progress towards proficiency. Accountability systems may also be informal, such as pressure exerted by publicity around test scores and school rankings. This might be operationalized by parents with increased pressure on school leaders and teachers, or by parents moving their children out of lower performing schools. It could also be enforced by principals, who have greater control over teacher hiring and firing than in traditional public schools.[2]

To describe potential score inflation in Boston charters, I draw heavily on the model used by Muralidharan and Sundararaman (2011), with some modifications. Teachers (who may be encouraged in a particular direction by their school leaders, both of whom are agents in this context), under the various formal and informal accountability systems described above, can spend time on two topics, T_1, frequently tested content, and T_2, infrequently tested content. In the context of this study, math and ELA would be considered frequently tested content, whereas science is infrequently tested. Within subjects, some curriculum standards are tested frequently and others are not (for details, see Section 2.4 on outcomes below). Additional time spent on frequently tested topics is represented by t_1 and additional time on infrequently tested topics is represented by t_2.

Both frequently and infrequently tested topics contribute to the production of gains in human capital:

$$H = f_1(t_1) + f_2(t_2) + \epsilon \tag{2.1}$$

where H is is unobserved gains in human capital, f_1 and f_2 are the marginal effects on human capital gains of time spent on t_1 and t_2, and ϵ is random error including all other factors that contribute to a student's gains in human capital. An education accountability system (the principal in the classic principal-agent problem) does not assign rewards and punishments to schools based

[2]Note that some accountability pressures are greater for charter schools than traditional public schools – reauthorization and teacher personnel decisions. However, this does not mean that I cannot compare the two types of schools, only that charter school leaders and teachers might face even more incentives to teach to the test.

on H, which is unobserved, but on an observable test score measure, Y. Test scores are also a function of time spent on frequently and infrequently tested content:

$$Y = g_1(t_1) + g_2(t_2) + \eta \tag{2.2}$$

where g_1 and g_2 are, respectively, the marginal effects of time spent on t_1 and t_2 on test scores and η is random error including all other factors that contribute to a student's test score. The key feature of this analysis is that the causal charter school effect, measured by exploiting the charter school lottery, can be broken into score subscales representing t_1 and t_2. Unlike a traditional principal-agent problem, an educational accountability system does not offer an explicit wage based on Y, but it offers school level rewards and punishments (which for charters, may include closure), perhaps consequences for individual teachers depending on how a school leader uses test scores (increased professional development, increased evaluation, more freedom, job security, termination), and psychological comfort from meeting accountability goals. These consequences do not directly affect salary or bonuses in most schools, but they do affect the non-pecuniary benefits of working in a school and can be considered part of a wage that is paid in utility.

Thus the accountability system offers a wage in utils, U, that is a function of the test score:

$$U = E[s] + E[Y] - E[C(t_1) + C(t_2)] \tag{2.3}$$

where $E[s]$ is the expected utility of the teacher's salary, $E[Y]$ the expected utility or disutility of the non-pecuniary benefits of test scores (note that $E[Y]$) may be negative) measured in dollars, and $E[C(t_1) + C(t_2)]$ is the expected utility of the costs associated with the effort of teaching. When trying to find an optimal contract, the next step in this model is to determine a bonus associated with Y that induces optimal behavior. Here, the above equations are sufficient to discuss how incentives from an accountability system may distort teacher behavior.

An increase (decrease) in test scores will increase (decrease) teacher utility. Additionally, if $g_1(t_1) > g_2(t_2)$ and $C(t_1) < C(t_2)$, reallocating time from infrequently tested items (T_2) to frequently tested items(T_1) will increase utility through two channels. First, when $g_1(t_1) > g_2(t_2)$ test scores will increase. Second, when $C(t_1) < C(t_2)$, costs will decrease. We expect $C(t_1) < C(t_2)$ if more curricular materials are provided for highly tested items and collaboration between teachers is easier for such items so that shifting time to t_1 lowers costs. Additionally, when T_1

is more emphasized on the test than T_2, it is likely that $g_1(t_1) > g_2(t_2)$ since additional t_1 will payoff on many items whereas additional t_2 will contribute to relatively few points on an exam. The most important question is whether f_1 has the same functional form as f_2 and both have non-decreasing returns. If both content areas influence gains in students' underlying human capital equally, it does not matter if teachers reallocate between T_1 and T_2. But if f_1 has decreasing marginal returns or if $f_2 > f_1$, reallocation to T_1 incentivized by the accountability system will lower human capital gains for students.

I argue that it is possible to separate Y into two components, Y_1 and Y_2, which in turn correspond to T_1 and T_2. For example, Y_1 measures performance on frequently tested content and Y_2 measures performance on infrequently tested content. I can then observe whether teachers respond to the incentive system that encourages them to increase Y by focusing on T_1, as measured by Y_1, or on T_2, as measured by Y_2.

Similar interpretations can be made if T_1 represents test preparation activities that increase Y but do not increase H (i.e. coaching) and T_1 represents other classroom activities that increase both Y and H.

2.4 Outcomes, Data, and Sample

2.4.1 Outcomes

Each of the outcome measures attempts to highlight a different way that instruction, and thus test scores, can be manipulated or reallocated. The outcome data come from detailed information from individual level MCAS results. Developed as a result of the 1993 Massachusetts Education Reform Act, which also allowed charters in the state, the MCAS has been the state's standardized test system since 1998. Since 2006, math and English/language arts have been tested in all of the relevant grade levels, and science is tested in 8th grade.

Using the detailed MCAS results, I added further information from the MCAS to create outcome variables that go beyond subject scores. Massachusetts makes public the question type, topic, difficulty, correct answer, and, since 2007, corresponding Massachusetts curriculum standard for each MCAS question (Massachusetts Department of Elementary and Secondary Education, 2011b).[3] Indeed, the state even publishes the actual MCAS question. Thus, when I merged these data with item level responses, I was able to identify each question that an individual student answered correctly and create outcome metrics based on subsets of questions.

The outcomes are grouped in three ways: rare standards, question type, and topic. Information on standards was first available for the spring 2007 MCAS, so outcomes using rarely tested standards have a restricted time range. I refer to this as the *rare standards sample*. Question type and question topic outcomes are available for all MCAS administrations, so I refer to these outcomes as covering the *full sample*. Each of the outcome measures is a standardized raw score of points in the category by subject, grade, and year. For reference, I also report outcomes for overall standardized score in each subject ("all items") in both the "rare standards" and "full" samples.

The MCAS exams consistently test each of the outcomes in similar proportions across years, making the frequently tested standards, question types, and topics on the test predictable. See Table 2.1 for details. For instance, in math, multiple choice items always account for about 30 points, short answer items about 5 points, and open response items about 19 points (the test format changed slightly in 2010). Topic areas also follow a consistent pattern across years.

The MCAS outcomes used here make up about 80% of the MCAS exam; the other 20% of the exam includes items for equating and trial purposes, which are not reported or included in score calculation but are similar in type and topic to the common 80% of items (Massachusetts Department of Elementary and Secondary Education, 2007). Thus schools and teachers can predict the format and topic of the MCAS each year. This predictability may lead to test preparation, as teachers can anticipate these features of each year's exam.

Rare Standards

For MCAS exams from spring 2007 to 2011, I determined which standards were given the most and least weight on the exams and divided the standards into terciles of rare standards, somewhat common standards, and common standards. This outcome allows me to assess whether charter school students do better on frequently assessed standards than on standards only assessed occasionally (to return to the theoretical model, T_1 and T_2). For instance, a question about

[3] Beginning in 2012, standards were categorized both by state standards and Common Core standards. Thus, I limit my sample to 2011 and prior years.

Table 2.1: *Average Points Possible on MCAS Items*

	Math			ELA			Science
	6th	7th	8th	6th	7th	8th	8th
Subscale Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Total Points Possible	54.0	54.0	54.0	52.0	52.0	52.0	54.0
	[11.8]	[12.1]	[12.6]	[8.3]	[8.6]	[8.6]	[10.1]
(A) Rare Standards Sample							
Rare	8.4	8.2	10.2	1.6	1.8	3.8	5.6
	[2.4]	[2.2]	[2.7]	[1.0]	[0.9]	[2.3]	[1.6]
Somewhat Common	19.6	13.0	9.8	9.8	7.4	9.4	13.2
	[4.6]	[3.2]	[2.7]	[2.3]	[2.2]	[3.5]	[3.0]
Common	26.0	32.8	34.0	40.6	42.8	38.8	35.2
	[6.2]	[7.6]	[8.4]	[6.6]	[7.0]	[7.7]	[6.8]
(B) Full Sample							
Multiple Choice	29.8	30.0	30.0	36.0	36.0	36.0	35.3
	[6.4]	[6.6]	[6.9]	[6.1]	[6.3]	[6.3]	[6.6]
Short Answer	5.3	5.3	5.3	-	-	-	-
	[1.5]	[1.6]	[1.7]				
Open Response	19.0	18.7	18.7	16.0	16.0	16.0	18.7
	[4.8]	[4.9]	[5.2]	[2.9]	[3.1]	[3.1]	[4.4]
Geometry	7.3	7.0	7.0	-	-	-	-
	[1.9]	[2.0]	[2.1]				
Measurement	7.1	7.0	7.0	-	-	-	-
	[2.2]	[2.1]	[2.3]				
Numbers & Operations	17.6	13.8	14.0	-	-	-	-
	[4.3]	[3.5]	[3.7]				
Patterns & Algebra	14.0	15.0	15.0	-	-	-	-
	[3.3]	[3.5]	[3.8]				
Statistics & Probability	8.0	11.2	11.0	-	-	-	-
	[2.2]	[2.8]	[2.7]				
Reading	-	-	-	45.6	47.2	46.0	-
				[7.4]	[7.9]	[7.7]	
Language & Literature	-	-	-	6.4	4.8	6.0	-
				[1.4]	[1.6]	[1.4]	
Earth & Space Science	-	-	-	-	-	-	13.5
							[2.9]
Life Science	-	-	-	-	-	-	14.0
							[3.0]
Physical Science	-	-	-	-	-	-	13.2
							[3.2]
Tech. & Engineering	-	-	-	-	-	-	13.3
							[2.9]

Notes: For the test years that contribute to these averages, see Table B.1. There is little variation across years. Statewide standard deviations are underneath points possible in brackets.

Massachusetts standard 8.N.11:

> Determine when an estimate rather than an exact answer is appropriate and apply in problem situations.

was asked only once between 2007 and 2011. In contrast, questions about Massachusetts standard 8.M.3:

> Demonstrate an understanding of the concepts and apply formulas and procedures for determining measures, including those of area and perimeter/circumference of parallelograms, trapezoids, and circles. Given the formulas, determine the surface area and volume of rectangular prisms, cylinders, and spheres. Use technology as appropriate.

were asked 21 times in 2007-2011, 5 times in 2007, 2008, and 2009, 4 times in 2010, and twice in 2011 (Massachusetts Department of Elementary and Secondary Education, 2000). While the second standard likely encompasses more concepts than the first standard, it is difficult to determine whether one or the other is more important for overall understanding of mathematics.

Question Type

Question type outcomes are multiple choice, short answer, or open response. Only the mathematics exams have short answer questions. Multiple choice questions and short answer questions are each worth one point on the exam and open response questions are worth four points, with students scoring zero to four on each open response. The format of question types was only changed once in the relevant period, with the math and science exams adding four multiple choice questions and subtracting one open response question in 2010. The format of the ELA exam was never changed in the relevant time period.

Topic

Question topic outcomes are specific to subject. For math they include geometry; measurement; number sense and operations; patterns, relations and algebra; and data analysis, statistics, and probability; for ELA they include reading and language and literature; and for science they include earth science; biology and life sciences, physical sciences, and technology and engineering. In math, number sense and operations and patterns, algebra, and relations are the most frequently tested topics, followed by data analysis, statistics, and probability. Geometry and measurement are tested the least in the middle school grades. In ELA, reading makes up the majority of the exam and language and literature items only make up a small portion of the test. Science topics are tested evenly. Across subjects, topic divisions are consistent across time. For instance, in ELA, reading accounts for 44 to 48 points on the exam and language and literature 4 to 8 points, depending on the test year.

2.4.2 Data

The data for this analysis come from statewide datasets provided by the Massachusetts Department of Elementary and Secondary Education, as well as lottery records collected from individual charter schools in Boston. The state provided data for school years 2001-2002 through 2009-2011 on students' demographic backgrounds, program participation, and school attendance, and MCAS scores in math, ELA, writing, and science. I assigned students to their most attended school in each year, except that students who attended at least one charter school were assigned to the charter school even if it was not their most attended school. Thus, a student who attended a charter school for one month and a student who attended a charter school for one year were both assigned to the charter school for that year. Since I attribute a full year of attendance and the students' tests scores to the charter schools, no matter how long the student attended, my results based on years of attendance can be considered a lower bound on the effect of attending a year of charter school.[4]

In addition to the state data, lottery records were collected from each charter school for the main entry grade in each school (5th or 6th grade). Lotteries were coded to identify students offered a seat at the charter school, to identify students who were never offered admission to the charter school, and to identify students that did not receive admissions offers randomly, such as students with sibling priority. Not all of the Boston area middle schools that admitted students for middle school entry in 5th or 6th grade were able to contribute records for lottery-based analysis. Two charter schools that contained middle school grades closed, two had insufficient

[4]Results where students are assigned to their most attended school, without an exception for charter schools, are quite similar. As predicted, these results are larger, but only by about 0.01–0.03σ indicating that my conservative assignment rule makes little difference in the conclusions of this study.

lottery records, and two admitted the majority of their students at the kindergarten level. Table B.10 includes details on school participation. The state data were combined with the lottery data through a matching process, which was then assembled into the analytic data set.

Since my focus is on middle school outcomes, I limit my dataset to students with baseline information from the grade of application to a charter (either 4th or 5th grade) who entered charter school lotteries in spring 2002 to spring 2010. The outcome scores available vary with subject and grade level and are detailed in Table B.1.

2.5 Methods

I estimate the causal effect of attendance at a charter school on student achievement in the same way as Abdulkadiroglu et al. (2009; 2011). However, since my intention is to disaggregate the charter school effect and determine if it is due to score inflation, the outcome measures are standardized components of the MCAS instead of subject scores, and are estimated separately by grade level, rather than pooled.

If all applicants who received an offer for a seat at a charter school attended that charter school and no applicants that did not receive an offer attended, that is, if all applicants were all compliers, OLS regression using a variable representing the receipt of an offer would be sufficient to estimate the effect of charter school attendance on outcomes. However, some applicants who received an offer to attend a charter school choose not to attend and a few students who lost the lottery ultimately attended a charter school,[5] I therefore use an instrumental variables approach to estimate the causal effect of charter school attendance on the outcomes of interest.

The causal effect of a year of charter school attendance on a test score outcome component is represented in Equation 2.4, the second stage of the instrumental variables estimation:

$$y_{it} = \alpha_t + \sum_j \delta_j d_{ij} + \beta' X_i + \rho S_{it} + \epsilon_{it}. \tag{2.4}$$

Here, y_{it} is the grade level specific test score based outcome of interest; S_{it} indicates the number of years of attendance, including repeated grades, at any charter school after the lottery at time t;

[5] These students likely were on the waitlist and were offered seats late in the school year or entered a lottery for a grade or obtained sibling preference subsequent to the entry year.

X_i is a vector of student level demographic and test score control variables determined before the lottery; and ϵ_{it} is an error term. I also include a set of year-of-outcome fixed effects, α_t, and a set of lottery fixed effects $\sum_j \delta_j d_{ij}$, that represent the charter school lottery risk set.[6]

Since attendance at a charter school is not randomly assigned, I use the charter school lottery offer, which is randomly assigned, as an instrument for years of charter school attendance.[7] In Equation 2.5, I represent the first stage:

$$S_{it} = \gamma_t + \sum_j \lambda_j d_{ij} + \kappa' X_i + \pi Z_{it} + \eta_{it}. \tag{2.5}$$

Here, S_{it} is estimated by X_i, a vector of student baseline demographic and test score control variables; $\sum_j \lambda_j d_{ij}$, a set of lottery fixed effects that represent the charter school lottery risk set; γ_t a set of year-of-outcome fixed effects; η_{it}, an error term; and the instrument, Z_i, which is a dummy variable that indicates if a charter school lottery applicant has received an offer to attend at least one charter school (sometimes referred to as winning the lottery).

In summary, π is the first stage effect, which in this case is the difference between the average number of years a student offered a seat at a charter school attends a charter school and the average number of years a student not offered a seat at a charter school attends a charter school. The causal effect of S_{it}, a year of charter school attendance, on y_{it}, the test score component, is ρ, which I also refer to as the local average treatment effect. The treatment effect is local since it applies only to compliers, and since it is estimated using a partial compliance estimator, it can also be referred to as the average causal effect (Angrist and Imbens, 1995). The associated reduced form or intent-to-treat effect, or effect of Z_i on y_{it}, is found in an equation similar to Equation 2.4 where Z_i is substituted for S_{it}. The coefficient of interest is ρ, which is the causal effect of a year of charter attendance, and is the ratio of the reduced form coefficient (difference in test based outcome between those offered a seat and those not offered a seat) to the first stage coefficient

[6]The charter school lottery risk set for any given applicant is a dummy variable representing the charter school entry grade lottery or lotteries that the applicant has applied to. For instance, applicants applying only to charter school A would be a in one risk set, applicants applying only to charter school B would be in another risk set, and applicants applying to both charter schools A and B would be in a third risk set. In Massachusetts, each charter runs its lottery independently, and students can apply to multiple charter schools. Since I only include lotteries for the main entry grades at schools, risk sets do not include later or repeat applications.

[7]I exclude siblings, since they are guaranteed admission to charter schools. I also exclude late applicants and applicants from out-of-area, who are sent to the bottom of the waitlist. I also verify the lottery by comparing pretreatment covariates in Table B.2, finding in a joint F test that there is no difference between the groups.

(difference in years of attendance at a charter school between those offered a seat and those not offered a seat).

2.6 Results

I fit the 2SLS model described above for each of my MCAS outcomes, such as math multiple choice score and science rare standards score.[8] I then inspect these outcomes to determine the composition of the middle school effects and its consistency or inconsistency across sections of the MCAS. I can also look for the effects of differential test preparation.

By comparing treatment effects across MCAS outcomes, I can see if the treatment effect for one or more of the outcomes has a larger response than the treatment effect on other outcome types. For the question type outcomes (multiple choice, short answer, and open response), differential success across outcomes may indicate that charter schools have coached to that question type to a greater extent than the other Boston schools attended by charter lottery losers. Likewise, by comparing treatment effects across standard frequency and subject topics, I can observe if results by standards frequency and topic are substantially different from each other. If charter school students are much more successful on common standards rather than typical standards, or certain frequently tested math topics rather than others, I would have evidence that charter schools are reallocating effort to teaching certain math standards and topics to a greater extent that other schools in Boston.

There are two important caveats. First, if charter schools are using coaching or reallocation with the same frequency as BPS, I expect to see no difference in treatment effects due to coaching or reallocation. For instance, if both charter schools and other public schools are teaching students guessing strategies for multiple choice items, the subscore for multiple choice items would not stand out, even if test preparation occurred. Additionally, if charter schools are effective at coaching across *all* types of test questions, or are reallocating from untested subjects to all tested standards and topics, then I could not identify a coaching or reallocation effect, since no

[8]Throughout this paper, I control for both baseline demographic characteristics and baseline test scores, which reduces the sample slightly. I focus on this model since it is the preferred model in prior work on Boston charters. Results are similar for a model that does not control for demographics or test scores and one that only controls for baseline demographics.

outcome would stand out. However, if charters are coaching a particular item type more than the comparison schools and more than other item types, I would expect to see a differential treatment effect on that item type subscale. Similarly, if charter schools are reallocating to common standards or more highly tested topics within a subject, I would expect to see higher scores on the more frequently tested items and lower scores on the less frequently tested standards and topics.

2.6.1 First Stage

In Table 2.2 I present the first stage results that show that the offer of a seat at a charter school does predict future attendance at charter schools. Results are similar across samples and subjects. By 6th grade, on average, students who are offered a seat through the charter lottery attend about 0.6 years more of charter school than students who did not receive an offer of a seat. By 7th grade, on average, students who are offered a seat through the charter lottery attend a charter for a full year more than students that did not receive an offer. By 8th grade, the difference is over a year and a half.

The first stage estimate may be less than the total potential time a student could attend a charter for two reasons. First, only 70% of students who win the lottery at one of the oversubscribed charter schools attend a charter school. Second, a third of the students who did not win a seat through an oversubscribed lottery nonetheless attended some charter school for some time. These latter students could attend a charter by entering at a later grade, obtaining sibling preference, getting a spot off the waitlist late in the school year, or attending a charter not included in the lottery sample.

2.6.2 Reduced Form and 2SLS

In Table 2.3, I present the reduced form results for rare standards, science, and question type. These results show the effect of being offered a seat at an oversubscribed charter school on MCAS subscale outcomes. Recall that the outcomes are standardized subscores, so that a statistically significant reduced form effect can be interpreted as the additional standard deviations (σ) correct on the MCAS subscore that a student offered a seat at a charter school scores compared to students not offered a seat. I present results separately by grade level since different grades test particular

Table 2.2: *First Stage Effect of a Lottery Offer on Years of Attendance at a Charter School*

	Math			ELA			Science
	Grade 6	Grade 7	Grade 8	Grade 6	Grade 7	Grade 8	Grade 8
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(A) Rare Standards Sample							
Years in Charter	0.718***	1.018***	1.395***	0.717***	1.021***	1.397***	1.398***
	(0.071)	(0.090)	(0.141)	(0.070)	(0.089)	(0.140)	(0.141)
N	2683	2194	1756	2677	2172	1751	1755
(B) Full Sample							
Years in Charter	0.699***	1.039***	1.402***	0.706***	1.049***	1.402***	1.405***
	(0.061)	(0.085)	(0.134)	(0.066)	(0.080)	(0.134)	(0.135)
N	3317	2373	1891	2987	2488	1889	1890

Notes: This table reports coefficients on regressions predicting years spent in a charter using the offer of a enrollment at a charter school. Each outcome cell is estimated by a separate regression. All regressions include baseline demographic controls, baseline test score controls, lottery risk sets, which are a set of dummies for the combination of schools applied to by year, and year of test and year of birth dummies. The sample is restricted to charter school applicants without sibling priority in the lottery, who attended a public or charter charter school in their year of application, and who have baseline demographic characteristics. Regressions use robust standard errors and are clustered by school by year. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

topics with varying frequency, but include pooled results for non-topic specific subscales in Table B.3.[9]

The causal effects of attending a year of charter school on MCAS outcomes are simply the ratio of the reduced form coefficients in Table 2.3 to the first stage coefficients in Table 2.2. In Table 2.4 , I show the 2SLS results, or average causal effects per year of attendance at a charter on MCAS subscale outcomes. Since the causal effects are "per year" of charter school attendance, the intention-to-treat effects in Table 2.3 will be scaled up or down to be equivalent to a year of charter school. Thus the 6th grade 2SLS effects are larger than those reported in Table 2.3 . The 7th grade 2SLS effects are about the same, and the 8th grade 2SLS results are smaller than the corresponding reduced form results.

In Table B.4. I also report the mean outcome score for lottery applicants who did not win a seat in the lottery and those that did in raw MCAS score points. The difference between the two means are roughly equivalent to the reduced form estimates. (The reduced form estimates also include control variables to increase statistical precision.) The mean scores gives context to the causal effects that report in the tables in standard deviation units. On the overall scores, students offered a seat in the lottery tend to outscore their counterparts not offered a seat by 3.5-4 MCAS raw points in math, 0.5-2 MCAS raw score points in ELA, 3-4 raw score points in science, depending on the particular sample. In ELA, the difference is only one multiple choice item on the test, but in math and science the difference is as large as 3-4 multiple choice items or the full score on an open response item. Since these overall gaps are spread across multiple subscales, and some subscales are only a few MCAS points themselves, differences in raw score points between offered and non-offered students will be smaller.

Rare vs. Common Standards

To examine whether charter schools are reallocating more than public schools from less frequently tested topics within each subject, Table 2.3 presents results for the reduced form and Table 2.4 for the 2SLS results of the charter school impact on rarely tested standards, somewhat common

[9]When results are not grade specific, pooled results show similar findings to the disaggregated results.

Table 2.3: *Reduced Form Effect of a Lottery Offer on MCAS Outcomes*

Subscale Outcome	Math Grade 6 (1)	Math Grade 7 (2)	Math Grade 8 (3)	ELA Grade 6 (4)	ELA Grade 7 (5)	ELA Grade 8 (6)	Science Grade 8 (7)
(A) Rare Standards Sample							
All Items	0.340***	0.303***	0.348***	0.139***	0.242***	0.171***	0.410***
	(0.038)	(0.049)	(0.056)	(0.032)	(0.043)	(0.051)	(0.063)
Rare	0.404***	0.367***	0.280***	0.182***	0.116**	0.090	0.279***
	(0.046)	(0.059)	(0.060)	(0.060)	(0.051)	(0.059)	(0.060)
Somewhat Common	0.393***	0.286***	0.293***	0.148***	0.191***	0.151***	0.259***
	(0.044)	(0.048)	(0.058)	(0.037)	(0.046)	(0.057)	(0.061)
Common	0.243***	0.258***	0.355***	0.116***	0.241***	0.178***	0.440***
	(0.035)	(0.047)	(0.056)	(0.034)	(0.044)	(0.052)	(0.066)
N	2683	2194	1756	2677	2172	1751	1755
(B) Full Sample							
All Items	0.369***	0.326***	0.385***	0.122***	0.233***	0.183***	0.418***
	(0.036)	(0.046)	(0.057)	(0.031)	(0.039)	(0.048)	(0.060)
Multiple Choice	0.383***	0.356***	0.380***	0.129***	0.205***	0.157***	0.416***
	(0.038)	(0.050)	(0.054)	(0.030)	(0.039)	(0.045)	(0.063)
Short Answer	0.377***	0.309***	0.354***	-	-	-	-
	(0.043)	(0.053)	(0.063)	-	-	-	-
Open Response	0.278***	0.234***	0.342***	0.068	0.208***	0.190***	0.351***
	(0.035)	(0.042)	(0.060)	(0.044)	(0.050)	(0.067)	(0.057)
N	3317	2373	1891	2987	2488	1889	1890

Notes: This table reports coefficients on regressions predicting MCAS outcomes using the offer of a enrollment at a charter school. Each outcome cell is estimated by a separate regression, using subscales standardized in the statewide sample by subscale and grade. All regressions include baseline demographic controls, baseline test score controls, lottery risk sets, which are a set of dummies for the combination of schools applied to by year, and year of test and year of birth dummies. The sample is restricted to charter school applicants without sibling priority in the lottery, who attended a public or charter charter school in their year of application, and who have baseline demographic characteristics. Regressions use robust standard errors and are clustered by school by year. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 2.4: *2SLS Effect of Attending a Charter School, Per Year of Attendance, on MCAS Outcomes*

Subscale Outcome	Math Grade 6 (1)	Grade 7 (2)	Grade 8 (3)	ELA Grade 6 (4)	Grade 7 (5)	Grade 8 (6)	Science Grade 8 (7)
(A) Rare Standards Sample							
All Items	0.474***	0.298***	0.250***	0.194***	0.237***	0.122***	0.293***
	(0.048)	(0.039)	(0.033)	(0.043)	(0.038)	(0.032)	(0.037)
Rare	0.563***	0.360***	0.200***	0.236***	0.113**	0.064	0.200***
	(0.061)	(0.051)	(0.035)	(0.083)	(0.048)	(0.040)	(0.039)
Somewhat Common	0.548***	0.281***	0.210***	0.206***	0.187***	0.108***	0.185***
	(0.058)	(0.037)	(0.033)	(0.050)	(0.043)	(0.037)	(0.041)
Common	0.339***	0.254***	0.254***	0.162***	0.236***	0.127***	0.314***
	(0.044)	(0.039)	(0.035)	(0.045)	(0.039)	(0.032)	(0.038)
N	2683	2194	1756	2677	2172	1751	1755
(B) Full Sample							
All Items	0.528***	0.313***	0.274***	0.173***	0.222***	0.131***	0.297***
	(0.050)	(0.036)	(0.033)	(0.042)	(0.033)	(0.030)	(0.036)
Multiple Choice	0.548***	0.343***	0.271***	0.182***	0.195***	0.112***	0.296***
	(0.053)	(0.040)	(0.032)	(0.039)	(0.033)	(0.028)	(0.038)
Short Answer	0.540***	0.297***	0.252***	-	-	-	-
	(0.059)	(0.044)	(0.037)	-	-	-	-
Open Response	0.398***	0.226***	0.244***	0.096	0.198***	0.135***	0.250***
	(0.050)	(0.034)	(0.038)	(0.061)	(0.046)	(0.044)	(0.036)
N	3317	2373	1891	2987	2488	1889	1890

Notes: This table reports coefficients on regressions predicting MCAS outcomes using the offer of a enrollment at a charter school. Each outcome cell is estimated by a separate regression, using subscales standardized in the statewide sample by subscale and grade.. All regressions include baseline demographic controls, baseline test score controls, lottery risk sets, which are a set of dummies for the combination of schools applied to by year, and year of test and year of birth dummies. The sample is restricted to charter school applicants without sibling priority in the lottery, who attended a public or charter charter school in their year of application, and who have baseline demographic characteristics. Regressions use robust standard errors and are clustered by school by year. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

standards, and common standards.[10] There is no conclusive pattern. Within each subject, subscores are within 0.05σ to 0.2σ of each other. As a whole, results by standards are positive, significant, and fairly large for all but one subscale: rare items in 8th grade ELA. This single not significant results may be due to chance, given the large number of outcomes I am testing, or it may be due to some reallocation away from rare standards in 8th grade ELA. But as a whole, the pattern across the standards outcomes do not suggest a pattern of reallocation away from the least frequently tested items.

This setup assumes that each MCAS is a weighted random draw of items, with items weighted towards common standards, and that the 2007-2011 exams are similar in standards distribution to past exams. Teachers observe this over time and would have the opportunity to focus on the most common standards. However, perhaps teachers only focus on last year's exam and then reallocate their time away from untested standards. To test for this, I create variables indicating items with standards not on last year's test and items with standards on last year's test. This is only possible in 6th and 8th grade math and 8th grade science, as 7th grade math and all years of ELA standards are tested on every MCAS. The sample for this analysis is also limited to MCAS 2008-2011 administrations, since I need both item level standards data (2007-2011) and information about last year's exam (so 2007 cannot be included). I present results from this analysis in Table B.5. Again, there is no consistent pattern across subscales, with charter school students outperforming comparison students on both standards that were not tested in the previous year and on standards that were tested in the previous year.

To return to the theoretical framework outlined in Section 2.3, the content of commonly tested standards (or those on last year's test) would correspond to T_1 and the content related to rarely tested standards (or those not on last year's test) would correspond to T_2. I directly observe the MCAS scores related to this content, Y_1 and Y_2, respectively. Since the test score outcomes are of the same magnitude and significance level, I conclude that, in spite of incentives that may encourage differential test preparation, I do not have evidence of reallocation across standards.

[10]This sample is limited to MCAS years 2007-2011 since the state only began making item level information available in 2007. In 2012, the state began transitioning to Common Core standards, so I limit my period of examination to the time where data is available and there is one consistent set of standards.

Low vs. High Stakes Subjects

Above, I find no evidence of reallocation within subject content on the MCAS from frequently tested standards to less frequently tested standards. However, schools and teachers may not be reallocating their efforts within a subject, but rather, away from less tested subjects towards highly-tested subjects. Nationally, the Center on Education Policy reports school districts increasing instructional time on tested subjects and decreasing time on subjects like science, social studies, foreign languages, arts, and physical education since the implementation of NCLB (McMurrer, 2007). Although I cannot directly compare instructional time, I can investigate whether charter schools in Boston have similar impact on science as on math and ELA and, for the first time, present results on science for Boston charters.

While science is tested in Massachusetts, it is tested only once in grades 6 through 8 and results from the test do not enter the calculation of AYP during the study time period.[11] Similarly, they are not emphasized in the public presentation of results: each year the Boston Globe publishes proficiency MCAS rankings by district and schools. The science results are in a panel far below the math and ELA rankings (The Boston Globe, 2011). Since charters do not face the same accountability pressure for science results, they might reallocate their efforts away from science towards math and ELA. If so, I would expect the effect of winning the lottery (Table 2.3 , Column 7) and the average causal response of attending a charter school (Table 2.4 , Column 7) on science ”all items” scores to be much smaller in magnitude and potentially not significant. However, results for the 8th grade science MCAS are quite similar to the results for the 8th grade math MCAS. The 2SLS effect in the full sample is about a 0.25σ gain in math test scores and 0.29σ gain in science test scores, per year of attendance at a charter school. These gains are of similar size and are both significant at the 0.001 level. Thus, I find no conclusive evidence of reallocation away from science. Similar to the interpretation of the standards findings, my comparison of high vs. low stakes subjects is represented in the theoretical framework where T_1 corresponds to math and ELA and T_2 corresponds to science and I find similar test scores for each test type.

This finding is somewhat analogous to the findings from a recent evaluation of teacher incentives in India (Muralidharan and Sundararaman, 2011). Like the pressures in Massachusetts

[11]Masschusetts began including MCAS science scores in AYP calculations in 2012.

from NCLB, which incentivize math and ELA but not other subjects, in Muralidharan and Sundararaman's experiment teachers were explicitly rewarded for student achievement in math and reading, but not in science or social studies. However, authors found significant gains in all subjects, suggesting that teachers increase their efforts across all topics when they are facing incentives, that academic press on students transfers across subjects, or that there is spillover from highly-incentivized subjects to not incentivized subjects.

Multiple Choice vs. Open Response

The bottom panels of Table 2.3 and Table 2.4 present reduced form and 2SLS results, for all question types. Investigating question type should allow me to see evidence of coaching by question type. For instance, if charter schools were coaching a particular strategy on open response questions more than traditional schools did, I would expect to see a higher relative score for open response questions than for other question types. It is not entirely clear which question type would benefit the most from coaching. Multiple choice items can be coached with test taking techniques like the process of elimination or encouraging students to guess (since there is no penalty for guessing on the MCAS). Open response items can be coached by encouraging students to write down any answer, instead of leaving the response blank, or to use key words to signal structure. However, if there is differential coaching across question types, perhaps because it is easier to coach to one item type, it could appear with different effect sizes across question type. In this case, difficult to coach items would be represented by T_1 in the model and easy to coach items are represented by T_2.

In general, charter school students do just as well on each type of question as they do on the subject as a whole. For example, in 6th grade, the overall 2SLS effect on the math MCAS is 0.53σ and scores by question type are quite similar: multiple choice, 0.55σ; short answer, 0.54σ; and open response 0.40σ. In one case, 6th grade ELA, the 2SLS effect for one question type is not significant while there are significant results for the other question type: overall ELA gains of 0.17σ, multiple choice gains of 0.18σ and a not significant positive result of 0.10σ for open response. This exception may be due to chance (given the large number of outcomes I am examining, it's not surprising that one would not be significant), or it may be due to a lack of emphasis on writing in

6th grade. Either way, I still conclude that, for the most part, charter schools outperform their peers in traditional public schools on all question types and see no direct evidence of coaching to question type.[12]

Infrequently vs. Frequently Tested Topics

Reduced form results by MCAS topic are presented in Table 2.5; 2SLS results in Table 2.6. Examining content topics is a similar exercise to examining rarely tested standards. Some topics are consistently tested less frequently—geometry and measurement in math, language and literature in ELA. If charter students perform less well on less frequently tested content areas, I would have evidence of reallocation within subject to more highly tested content areas.

However, unlike students in Chicago, where the introduction of high-stakes testing resulted in differential effects by question topic (Jacob, 2005), charter school students do better than comparison students on all topics on the subject exams. While there is some fluctuation in the magnitude of effects across topics and grades, all show strongly significant positive results. Therefore, while I cannot rule out reallocation within math topics to those more frequently tested on the MCAS, I have no evidence of it. If both charter schools and the schools that charter lottery losers attend are reallocating their teaching efforts within the math exam to comparable extents, I also would not be able to detect evidence of reallocation.

2.7 Threats to Validity

2.7.1 Matching

Students offered a seat in a charter school lottery are more likely to be matched to the state database than students not offered a seat. This is likely due to lottery losers being more likely to enter private school. However, if these unmatched students are substantially higher performing than the matched lottery losers, their omission from the results would bias my findings upward.

[12] Another possibility is that charter school students have more interim assessments than their counterparts in traditional public schools and that this familiarity generates the success across all item types. I cannot directly test the number of interim assessments in the two sectors, as this is not reported in the data. However, BPS uses both required and teacher generated formative assessments through Assessment Technology Incorporated (ATI), which exposes students to standardized testing in the traditional public school setting as well.

Table 2.5: *Reduced Form Effect of a Lottery Offer on MCAS Topics*

		Math			ELA		Science
Subscale Outcome	Grade 6 (1)	Grade 7 (2)	Grade 8 (3)	Grade 6 (4)	Grade 7 (5)	Grade 8 (6)	Grade 8 (7)
(A) Full Sample							
Geometry	0.375***	0.325***	0.366***	-	-	-	-
	(0.042)	(0.046)	(0.068)	-	-	-	-
Measurement	0.393***	0.351***	0.327***	-	-	-	-
	(0.038)	(0.050)	(0.056)	-	-	-	-
Number Sense & Operations	0.257***	0.197***	0.334***	-	-	-	-
	(0.042)	(0.041)	(0.056)	-	-	-	-
Patterns, Algebra & Relations	0.280***	0.263***	0.320***	-	-	-	-
	(0.034)	(0.050)	(0.053)	-	-	-	-
Data Analysis, Statistics & Probability	0.282***	0.278***	0.375***	-	-	-	-
	(0.037)	(0.048)	(0.058)	-	-	-	-
Reading	-	-	-	0.197***	0.183***	0.128**	-
	-	-	-	(0.040)	(0.041)	(0.050)	-
Language and Literature	-	-	-	0.098***	0.226***	0.183***	-
	-	-	-	(0.032)	(0.040)	(0.050)	-
Earth and Space Science	-	-	-	-	-	-	0.309***
	-	-	-	-	-	-	(0.062)
Life Science	-	-	-	-	-	-	0.432***
	-	-	-	-	-	-	(0.062)
Physical Science	-	-	-	-	-	-	0.453***
	-	-	-	-	-	-	(0.063)
Technology and Engineering	-	-	-	-	-	-	0.240***
	-	-	-	-	-	-	(0.054)
N	3317	2373	1891	2987	2488	1889	1890

Notes: The notes for this table is the same as the notes for Table 2.3, only the outcomes differ. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 2.6: *2SLS Effect of Attending a Charter School, Per Year of Attendance, on MCAS Topics*

	Math			ELA			Science
	Grade 6	Grade 7	Grade 8	Grade 6	Grade 7	Grade 8	Grade 8
Subscale Outcome	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(A) Full Sample							
Geometry	0.537***	0.313***	0.261***	-	-	-	-
	(0.062)	(0.038)	(0.041)	-	-	-	-
Measurement	0.562***	0.338***	0.233***	-	-	-	-
	(0.055)	(0.042)	(0.033)	-	-	-	-
Number Sense & Operations	0.367***	0.190***	0.238***	-	-	-	-
	(0.061)	(0.036)	(0.035)	-	-	-	-
Patterns, Algebra & Relations	0.401***	0.253***	0.228***	-	-	-	-
	(0.045)	(0.040)	(0.033)	-	-	-	-
Data Analysis, Statistics & Probability	0.403***	0.268***	0.267***	-	-	-	
	(0.049)	(0.038)	(0.037)	-	-	-	
Reading	-	-	-	0.279***	0.175***	0.091***	-
	-	-	-	(0.058)	(0.039)	(0.034)	-
Language and Literature	-	-	-	0.138***	0.215***	0.131***	-
	-	-	-	(0.042)	(0.035)	(0.031)	-
Earth and Space Science	-	-	-	-	-	-	0.220***
	-	-	-	-	-	-	(0.038)
Life Science	-	-	-	-	-	-	0.307***
	-	-	-	-	-	-	(0.038)
Physical Science	-	-	-	-	-	-	0.322***
	-	-	-	-	-	-	(0.039)
Technology and Engineering	-	-	-	-	-	-	0.171***
	-	-	-	-	-	-	(0.036)
N	3317	2373	1891	2987	2488	1889	1890

Notes: The notes for this table is the same as the notes for Table 2.4, only the outcomes differ. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

To address this possibility, I present results in Table B.7 that included only applicants from the 2002 and 2009 spring lotteries, which do not have a significant difference in match rates between the offered and non offered groups (Table B.6). I only show 6th grade results because of small sample sizes for the higher grades. While there is some volatility in the results, as a whole they are just as large or even larger than the findings for the full sample, leading me to conclude that differential match rates are not biasing the results.

2.7.2 Attrition

If students leave the sample at different rates based on their offer or lack of an offer of a seat at a charter school, the results may be biased if students who leave differ in unobserved ways from students who say. Table 2.7 shows that there is no significant differential attrition between students offered and not offered a seat. However, in case there are unobserved patterns among attriters that could influence outcomes, I refit my results including attriters, by using baseline test scores as substitutes for missing middle grade outcomes (baseline math score is used for all math and science outcomes, baseline ELA score for ELA outcomes). This model assumes that students with missing outcomes continue to perform at the same level as at baseline. In actuality, performance at the exact same level between baseline grade and middle school is unlikely, but it is a good proxy since test scores are strongly correlated across grades ($r \approx .75$). With baseline scores assigned for missing outcomes, the findings are essentially the same as those presented in Section 2.7 (Table B.8, for brevity I present only the 2SLS results). Since there is little to no difference between the original findings and the results with baseline test scores assigned to missing outcomes, I conclude that the findings are not biased by selective attrition.

2.7.3 Reallocation between Students

Instead of reallocating resources to highly tested areas in order to boost scores, charter schools may be reallocating resources to particular students to increase test scores. Focusing on students for whom intervention is mostly likely to influence proficiency categorization could increase test scores due to differential treatment effects by student type. Several studies have found that schools and teachers focus on students who are on the verge of proficiency (which is the test score

Table 2.7: *Attrition*

	ELA		Math		Science	
	Proportion of Non-Offered with MCAS	Difference	Proportion of Non--Offered with MCAS	Difference	Proportion of Non--Offered with MCAS	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Has 6th Grade Outcomes	0.874	-0.004	0.875	-0.006	-	-
		(0.007)		(0.007)		-
N	1494	3410	1332	3052	-	-
Has 7th Grade Outcomes	0.880	0.012	0.875	0.013	-	-
		(0.008)		(0.009)		-
N	915	2396	1014	2544	-	-
Has 8th Grade Outcomes	0.859	0.002	0.861	-0.001	0.859	0.001
		(0.009)		(0.009)		(0.009)
N	752	1920	705	1913	752	1920

Notes: This table reports coefficients on regressions of an indicator variable equal to one if the outcome test score is non-missing on an indicator variable equal to one if the student was offered a seat in the lottery. The regressions are separate for grade level of outcome. All regressions include baseline demographic controls, baseline test controls, lottery risk sets, which are a set of dummies for the combination of schools applied to by year, and year of birth dummies. The sample is restricted to charter school applicants without sibling priority in the lottery, who attended a public or charter charter school in their year of application, and who have baseline demographic characteristics. Standard errors are robust. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

outcome used in AYP calculations), perhaps to the detriment of other students. In Chicago, Neal and Schanzenbach (2010) show differential test score increases for students in the middle of the test score distribution, the so-called "bubble kids," and a case study from Texas demonstrates this is an explicit pattern in some schools (Booher-Jennings, 2005).

In order to determine if charter schools are focusing on students on the verge of or just above proficiency to a greater degree than their traditional school counterparts, I include interaction terms in the model that estimate the effect of charter school attendance for students within four scaled score points of the baseline proficiency threshold in the baseline grade. For example, the proficiency threshold is 240, so students scoring 236 and 238 are considered near and underneath the threshold in their baseline year, and students scoring 240 and 242 are considered near and above the threshold in their baseline year.[13] This baseline definition attempts to both measure prior proficiency levels in the way a school or teacher would when examine the records of individual students, and also to avoid concerns about endogeneity. I present interaction results only for 6th grade outcomes, since these are the closest to when prior proficiency is determined.[14]

Since Massachusetts AYP determinations are based on a state calculated Composite Performance Index (CPI) that also gives credit to some scores below proficiency, I also create "near" variables for each kink in the CPI calculation. CPI points are awarded as such: proficient or above (above 240 MCAS points), 100 CPI points; needs improvement high (230-238 MCAS points), 75 CPI points; needs improvement low (220-228 MCAS points), 50 CPI points; warn/fail high (210-218 MCAS points), 25 CPI points; and warn/fail low (200-208 MCAS points), 0 CPI points. Massachusetts also allows schools to achieve AYP through improvement, which involves a specific goal set for each school and subgroup. However, improvement is also calculated using the CPI, with its kinked nature, which would again put the focus on students near thresholds rather than throughout the achievement distribution.

I investigate the interaction between years of attendance at a charter school and prior overall standardized score (Table 2.8 for math and Table 2.9 for ELA). To test whether overall score or specific place in the score distribution is relevant, I do this both for overall standardized score,

[13]The MCAS is scored in multiples of two, rangling from 200–280.

[14]Results (not shown) are similar in 7th and 8th grade.

and in a separate models, near each prior CPI relevant threshold: proficient, needs improvement high, needs improvement low, and warn/fail high. If charter schools are focusing on students on the "bubble" of proficiency (or another score threshold) to a larger extent than their traditional public school counterparts, I would expect the interaction terms for students in the prior year near the threshold category to have a significant positive contribution to the test score impacts (Columns 4, 6, 8, and 10). However, this is the case for none of the math outcomes and only one of the ELA outcomes (perhaps, given the large number of coefficients tested, due to chance). Instead, it appears that the charter school effect is largest across all math outcomes and two of the ELA outcomes for students with the lowest prior test scores (Column 2). Thus I find little evidence in test score outcomes that charters are focusing on students on the verge of proficiency or another score threshold at a rate greater than the schools that their counterparts attend. The charter schools are in fact most effective, at least in math, for the many students at the very bottom of the proficiency distribution.

2.8 Conclusion

This paper investigates the details of the large causal impacts of attendance on MCAS outcomes at highly-demanded middle school charters in Boston. Despite an incentive structure that would seem to reward teachers and charter schools for focusing on certain aspects of MCAS tests, I find no evidence of test preparation in comparison to traditional public schools. The consistent results across all elements of the test provide no discernible evidence of more reallocation between rare and common standards, low and high stakes subjects, multiple choice and open response questions, and infrequently and frequently tested topics in charter schools compared to traditional public schools. These results remain substantively the same when baseline test scores are assigned to those with missing outcomes or when limited to the sample with the same match rate by offer status. Nor is there evidence that charter schools are focusing on "bubble" students at a greater rate than other schools in Boston. My analysis strategy cannot conclusively rule out inappropriate test preparation, especially if it is consistent across all aspects of the test or if it is comparable to the test preparation that comparison schools conduct. However, the evidence

Table 2.8: *2SLS with Interactions: Math*

Subscale Outcome	Prior Score Main Effect (1)	Prior Score Interaction (2)	Prior Near Proficient Threshold Main Effect (3)	Prior Near Proficient Threshold Interaction (4)	Prior Near NI High Threshold Main Effect (5)	Prior Near NI High Threshold Interaction (6)	Prior Near NI Low Threshold Main Effect (7)	Prior Near NI Low Threshold Interaction (8)	Prior Near Warn High Threshold Main Effect (9)	Prior Near Warn High Threshold Interaction (10)
(A) Rare Standards Sample										
All Items	0.454***	-0.156***	0.488***	-0.093**	0.480***	-0.060	0.475***	-0.008	0.471***	0.244
	(0.046)	(0.042)	(0.050)	(0.043)	(0.048)	(0.041)	(0.049)	(0.054)	(0.048)	(0.163)
Rare	0.406***	-0.097**	0.574***	-0.079	0.563***	-0.009	0.562***	0.007	0.562***	0.092
	(0.056)	(0.046)	(0.062)	(0.052)	(0.062)	(0.040)	(0.062)	(0.056)	(0.061)	(0.154)
Somewhat Common	0.528***	-0.104*	0.564***	-0.111**	0.550***	-0.020	0.549***	-0.014	0.545***	0.247
	(0.068)	(0.057)	(0.060)	(0.043)	(0.059)	(0.046)	(0.059)	(0.065)	(0.058)	(0.174)
Common	0.417***	-0.158***	0.349***	-0.075	0.347***	-0.087*	0.339***	-0.008	0.336***	0.254*
	(0.044)	(0.041)	(0.046)	(0.047)	(0.044)	(0.045)	(0.044)	(0.052)	(0.044)	(0.153)
(B) Full Sample										
All Items	0.497***	-0.174***	0.545***	-0.130***	0.536***	-0.074*	0.530***	-0.017	0.527***	0.174
	(0.046)	(0.043)	(0.052)	(0.043)	(0.050)	(0.039)	(0.050)	(0.051)	(0.050)	(0.126)
Multiple Choice	0.512***	-0.197***	0.562***	-0.103**	0.555***	-0.068	0.548***	0.003	0.547***	0.139
	(0.048)	(0.045)	(0.054)	(0.044)	(0.053)	(0.044)	(0.053)	(0.057)	(0.053)	(0.125)
Short Answer	0.514***	-0.142***	0.553***	-0.103*	0.551***	-0.110**	0.539***	0.003	0.538***	0.190
	(0.056)	(0.053)	(0.062)	(0.057)	(0.061)	(0.046)	(0.060)	(0.066)	(0.059)	(0.165)
Open Response	0.376***	-0.121***	0.418***	-0.155***	0.404***	-0.057	0.402***	-0.041	0.396***	0.193
	(0.048)	(0.043)	(0.051)	(0.048)	(0.049)	(0.043)	(0.050)	(0.055)	(0.050)	(0.125)

Notes: This table reports coefficients on regressions predicting math test-score based outcomes using years spent in charter school and an interaction between years spent in a charter school and a prior test outcome as predicted by the offer of a enrollment at a charter school and the offer interacted with the prior test outcome. The remaining notes are the same as those for Table 2.4. Sample sizes are the same as those for 6th grade outcomes in Table 2.4. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 2.9: *2SLS with Interactions: ELA*

Subscale Outcome	Prior Score		Prior Near Proficient Threshold		Prior Near NI High Threshold		Prior Near NI Low Threshold		Prior Near Warn High Threshold	
	Main Effect (1)	Inter-action (2)	Main Effect (3)	Inter-action (4)	Main Effect (5)	Inter-action (6)	Main Effect (7)	Inter-action (8)	Main Effect (9)	Inter-action (10)
(A) Rare Standards Sample										
All Items	0.181***	-0.059	0.195***	-0.007	0.202***	-0.074	0.190***	0.072	0.191***	0.280
	(0.040)	(0.036)	(0.045)	(0.036)	(0.042)	(0.052)	(0.042)	(0.063)	(0.043)	(0.221)
Rare	0.231***	-0.022	0.261***	-0.143***	0.234***	0.017	0.235***	0.018	0.235***	0.060
	(0.047)	(0.044)	(0.053)	(0.052)	(0.050)	(0.063)	(0.049)	(0.091)	(0.050)	(0.362)
Somewhat Common	0.193***	-0.059	0.208***	-0.016	0.208***	-0.021	0.195***	0.196**	0.205***	0.103
	(0.047)	(0.044)	(0.053)	(0.052)	(0.050)	(0.063)	(0.049)	(0.091)	(0.050)	(0.362)
Common	0.150***	-0.055	0.160***	0.011	0.171***	-0.090*	0.160***	0.032	0.158***	0.322
	(0.042)	(0.038)	(0.047)	(0.034)	(0.044)	(0.053)	(0.044)	(0.060)	(0.045)	(0.196)
(B) Full Sample										
All Items	0.160***	-0.055	0.172***	0.005	0.182***	-0.083	0.171***	0.033	0.170***	0.211
	(0.039)	(0.035)	(0.044)	(0.036)	(0.041)	(0.051)	(0.041)	(0.064)	(0.042)	(0.179)
Multiple Choice	0.169***	-0.058*	0.185***	-0.018	0.188***	-0.050	0.181***	0.030	0.179***	0.277
	(0.038)	(0.034)	(0.041)	(0.039)	(0.039)	(0.056)	(0.039)	(0.072)	(0.040)	(0.191)
Open Response	0.514***	-0.142***	0.553***	-0.103*	0.551***	-0.110**	0.539***	0.003	0.538***	0.190
	(0.056)	(0.053)	(0.062)	(0.057)	(0.061)	(0.046)	(0.060)	(0.066)	(0.059)	(0.165)

Notes: This table reports coefficients on regressions predicting ELA test-score based outcomes using years spent in charter school and an interaction between years spent in a charter school and a prior test outcome as predicted by the offer of a enrollment at a charter school and the offer interacted with the prior test outcome. The remaining notes are the same as those for Table 2.4. Sample sizes are the same as those for 6th grade outcomes in Table 2.4. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

I show here also aligns with recent work showing that Boston charter high school[15] students outperform their counterparts on SAT and AP tests and are more likely to enroll in four-year colleges (Angrist, *et al.*, forthcoming)). Follow up work on the Harlem Children's Zone also finds positive outcomes on non state standardized test academic and social outcomes (Dobbie and Fryer, forthcoming). Combined with this recent evidence from the literature, the lack of any evidence of test preparation in these findings is suggestive that charter school gains are due to building the human capital of their students, rather than just increasing test scores, in spite of incentives that encourage teaching to the test.

[15]The sample overlap is quite small with the middle schools examined in this study, since few cohorts are currently old enough to observe these outcomes.

Chapter 3

Merit Aid, College Quality, and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy[1]

3.1 Introduction

Recent research has emphasized troubling trends in U.S. college completion rates over the past few decades. Among students entering college, completion rates are lower today than they were in the 1970s, due largely to low completion rates of men and students from lower socioeconomic backgrounds (Belley and Lochner 2007, Bailey and Dynarski 2011). This trend has spurred a vigorous debate over the relative importance of factors that vary across students, such as academic skill and family financial resources, and factors that vary across postsecondary institutions, such as funding levels or management quality. Distinguishing the influence of student-level and institution-level factors on college completion rates is confounded by the non-random selection of students into institutions of different apparent quality. In this paper, we provide further evidence consistent with the fact that the quality of the institutions themselves affects college completion rates.

[1]Co-authored with Joshua Goodman. This essay has also been published in the *American Economic Journal: Applied Economics* in 2014.

To do so, we exploit is a Massachusetts merit aid program in which high school students with test scores above multiple thresholds were granted tuition waivers at in-state public colleges. Such colleges were of lower quality than the average alternative available to such students, where quality is measured by a combination of graduation rates, academic skill of the student body, and instructional expenditures, as suggested by Black and Smith (2006). The scholarship, though relatively small in monetary value, induced substantial changes in college choice, allowing us to estimate the impact of college quality on students' postsecondary enrollment decisions and rates of degree completion. A regression discontinuity design comparing students just above and below the eligibility threshold finds that students are remarkably willing to forgo college quality for relatively little money and that marginal students lowered their college completion rates by using the scholarship. College completion rates decreased only for those subsets of students forgoing the opportunity to attend higher quality colleges when accepting the scholarship. We describe the magnitude of this response as remarkable because the value of the scholarship is dwarfed by estimates of the forgone earnings of attending a lower quality college or failing to graduate. As a whole, these results suggest that college quality has a substantial impact on college completion rates. We also find clear evidence that this scholarship increased college enrollment, though not graduation, rates for the most disadvantaged students. Such students comprise, however, an extremely small fraction of the total pool of those eligible for this merit aid.

Our research contributes to three strands in the literature on postsecondary education and the public subsidy of such education. First, a now extensive literature documents the sensitivity of students' college enrollment decisions to financial aid generally (Deming and Dynarski 2010, Kane 2006) and merit aid more specifically (Dynarski 2000, Cornwell et al. 2009, Dynarski 2008, Kane 2007, Pallais 2009, Goodman 2008[2]). In contrast to most of the programs studied in this literature, the Adams Scholarship targets a very highly skilled set of students, namely the top 25% of high school graduates in each school district. As a result, our estimates are generated by a part of the

[2]This paper represents an extension and improvement of Goodman (2008), which studied the same merit aid program at an earlier time and used less informative outcome data. In particular, that earlier paper could only measure whether graduating high school seniors self-reported their intention to enroll in public or private colleges, without identification of specific campuses, actual enrollment or persistence and graduation. This paper, in contrast, uses substantially more detailed administrative data that allows identification of the specific institutions students actually enroll in, as well measurement of persistence and graduation rates. This allows for clear estimation of the quality and cost tradeoffs students are making, the impact of this merit aid on in-state enrollment and, perhaps most importantly, the impact of this aid on college graduation rates.

skill distribution not often studied. Furthermore, unlike in most aid programs, recipients were automatically notified of their eligibility without having to apply. Simplifying the aid process is known to affect students' college enrollment decisions (Bettinger et al. 2012), so that this program design may explain in part the large impacts of aid observed here. Our results are also consistent with Fitzpatrick and Jones (2012), which finds that merit aid does effectively keep some students in state but that marginal students are a small fraction of total aid recipients.

Second, we add to the growing literature on the impact of college quality on student outcomes. Much of the literature on the impact of college quality on degree completion has focused on the community college sector, reaching varying conclusions about whether access to and quality of community colleges affects educational attainment (Rouse 1995, Leigh and Gill 2003, Sandy et al. 2006, Calcagno et al. 2008, Stange 2009, Reynolds 2012). Estimates of the impact of college quality on labor market earnings are similarly varied, with some positive (Loury and Garman 1995, Brewer et al. 1999, Chevalier and Conlon 2003, Black and Smith 2004, Black and Smith 2006, Long 2008, Hoekstra 2009, Andrews et al. 2012), some zero or positive only for disadvantaged sub-groups (Dale and Krueger 2002, Dale and Krueger 2011), and some suggesting that earnings differences dissipate once the job market properly understands graduates' underlying ability (Brand and Halaby 2006, Lang and Siniver 2011). Nearly all of these research designs attempt to eliminate selection bias either by conditioning on students' observable characteristics or by instrumenting college quality with distance from or tuition of nearby colleges. Neither approach entirely eliminates the possibility that unobserved student-level factors may be driving their estimates. The exception to this is Hoekstra (2009), which uses a discontinuity inherent in the admissions process to a flagship university to estimate the labor market return to an elite college education. We employ a similarly identification strategy and unlike Hoekstra are able to observe the college choice made by students not enrolling in the target institutions, allowing us to estimate the impact of merit aid on college quality. Though sources of exogenous variation in school and curriculum quality are more common at lower levels of schooling because of school choice lotteries (Deming et al. 2013) and test score-based admissions rules (Bui et al. 2014, Abdulkadiroğlu et al. 2014, Dobbie and Fryer 2014), they are rarer in the postsecondary literature.

Furthermore, our finding that college quality plays an important role in completion rates is consistent with important pieces of recent evidence. Controlling for rich sets of student characteristics does not eliminate wide variation among postsecondary institutions in completion rates (Bowen et al. 2009). Students who attend college in large cohorts within states have relatively low college completion rates, likely stemming from decreased resources per student given states' tendencies to change public postsecondary budgets slowly (Bound and Turner 2007). Bound et al. (2010) argue that the vast majority of the decline in completion rates can be statistically explained by decreasing resources per student within institutions over time and, even more importantly, shifts in enrollment toward the relatively poorly funded public sector. All of this suggests that characteristics of colleges themselves, such as resources available per student, play an important role in completion rates and that student-level factors are only part of the story.

Third, we show the empirical importance of the theoretical possibility first discussed in Peltzman (1973) that in-kind subsidies of public institutions can reduce consumption of the subsidized good. Prior work has shown how public in-kind subsidies can generate at least partial crowdout of privately provided health insurance (Cutler and Gruber 1996, Brown and Finkelstein 2008), preschools (Bassok et al. 2012) and two-year colleges (Cellini 2009). Peltzman's contribution was the prediction that, in some cases, crowdout could theoretically be large enough to reduce overall consumption of the subsidized good. Work by Ganderton (1992), using cross-state variation in tuition subsidies, and Long (2004), using much finer college-specific variation in such subsidies, suggests that this in-kind public support for postsecondary education does reduce overall spending on education. We contribute to this literature by providing the first evidence of such reduced consumption driven by an exogenous shock in the size of the in-kind subsidy. We also show that this reduced spending on higher education comes at the cost of a reduced probability of degree completion, a possibility recognized by Kane (2007) in his evaluation of the D.C. Tuition Assistance Grant program but unexplored because too little time had passed to look beyond enrollment effects.

The structure of the paper is as follows. In section 3.2, we describe the merit scholarship program in detail. In section 3.3, we describe the data on students and colleges, including our measures of college quality. In section 3.4, we explain our empirical strategy, a regression discontinuity design that accounts for the multiple thresholds students must cross in order to be eligible for aid. In section 3.5, we present estimates of the impact of college quality on enrollment decisions and completion rates. In section 3.6, we discuss implications of our findings and conclude in section 3.7.

3.2 The Adams Scholarship

All Massachusetts public high school 10th graders take the Massachusetts Comprehensive Assessment System (MCAS), which includes an English language arts (ELA) portion and a mathematics portion. Scores on each portion range in multiples of two from 200 to 280, with 260-280 categorized as "advanced" and 240-258 as "proficient". In January 2004, Massachusetts Governor Mitt Romney proposed the John and Abigail Adams Scholarship Program, which would waive tuition at in-state public colleges for any student whose total MCAS score placed him or her in the top 25% of students statewide.[3] Romney's two stated goals seemed to be keeping highly talented students in state and improving the quality of the state's public postsecondary institutions. In his January 15, 2004 State of the Commonwealth speech to the Massachusetts legislature, Governor Romney explained that "I want our best and brightest to stay right here in Massachusetts."[4] Conversations with individuals involved with the scholarship's inception also suggest that Massachusetts wanted the recently introduced MCAS exam to be seen as a valid measure of student achievement and was thus willing to, in effect, put its money where its mouth was.

Concerned that Governor Romney's statewide standard would assign scholarships largely to students in wealthy, high-performing school districts, the state Board of Higher Education ultimately approved a modified version of the program in October 2004.[5] Under the approved policy, which has continued through at least 2013, a student receives a tuition waiver if his or her

[3]The eponymous couple cared deeply about education. John Adams wrote, in the Massachusetts Constitution, that "Wisdom, and knowledge, as well as virtue... as these depend on spreading the opportunities and advantages of education in the various parts of the country, and among the different orders of the people, it shall be the duty of legislatures and magistrates, in all future periods of this commonwealth, to cherish the interests of literature and the sciences, and all seminaries of them; especially the university at Cambridge, public schools and grammar schools in the towns" (Chapter V, Section II). Abigail Adams, disturbed by the 18th century gender gap, wrote that "It is really mortifying, sir, when a woman possessed of a common share of understanding considers the difference of education between the male and female sex, even in those families where education is attended to" (Letter to John Thaxter, February 15, 1778).

[4]See the January 20, 2004 Boston Globe article, "Specialists Blast Romney Proposal for Free Tuition," by Jenna Russell.

[5]See the October 20, 2004 Boston Globe article, "New MCAS Scholarship OK'd," by Jenna Russell.

MCAS scores fulfill three criteria. First, he or she must score advanced on one portion of the exam. Second, he or she must score proficient or advanced on the other portion of the exam. Third, the student's total MCAS score must fall in the top 25% of scores in his or her school district.[6] The scores used to determine eligibility come from each student's first attempt at taking the grade 10 MCAS tests in ELA and mathematics. To receive the scholarship, a student must be enrolled in and graduate from a Massachusetts public high school in his or her senior year. The graduating class of 2005 was the first to receive Adams scholarships.

Scholarship winners are automatically notified by letter in the fall of their senior year. The scholarship waives tuition at any of four University of Massachusetts (U. Mass.) campuses, nine (four-year) state colleges, or fifteen (two-year) community colleges.[7] As such, the letter that Governor Romney sent to the first class of scholarship recipients promised in bold-faced and underlined letters "four years of free tuition." Receipt of the scholarship does not, however, eliminate the cost of college attendance. To clarify the distinction between tuition and fees, the letter to the second class of scholarship recipients added to its final paragraph the disclaimer that "College fees and rooming costs are not included in this scholarship award." More recent letters have emphasized this fact even more clearly.[8]

Figure C.4 shows the tuition and mandatory fees at the University of Massachusetts at Amherst and Bridgewater State College, the two largest campuses in their respective sectors. Strikingly, at both campuses and nearly all other public Massachusetts colleges, tuition has remained constant in nominal terms over the past decade. Mandatory fees have, however, risen dramatically.[9] For the first class of scholarship winners in 2005, the tuition waiver was worth $1,714 annually if used at U. Mass. Amherst or $910 if used at Bridgewater State. Given mandatory fees of $7,566 at U. Mass. Amherst and $4,596 at Bridgewater State, the Adams Scholarship thus represented a roughly 20%

---

[6]As of the class of 2006, students in charter schools or who participate in school choice or the Metco program can fulfill the third criterion by placing in the top 25% of the district they attend or the district in which they reside.

[7]Six of Massachusetts' state colleges (Salem, Bridgewater, Fitchburg, Framingham, Westfield and Worcester) were renamed "state universities" in 2010. For simplicity, we refer to them as "state colleges" throughout the paper.

[8]See Figures C.1, C.2 and C.3 for copies of these letters.

[9]This peculiar detail may be due to the fact that tuitions are set by the Massachusetts Board of Higher Education and flow directly to the state's General Fund, while fees are set by each college's Board of Trustees and are retained by the colleges themselves.

reduction in the direct cost of attendance. By the fall of 2011, fees had risen by more than a third, so that the Adams Scholarship represented a less than 15% reduction in the cost of attendance. These percentages would be substantially lower if room, board and other expenses were included in the total cost of attendance. Conversations with individual colleges' financial aid offices also suggest that for some students this aid is factored into financial aid offers and may be partially crowded out as a result.[10] The Adams Scholarship thus lowers the cost of college attendance by well under 20%, may be partially crowded out by college financial aid offices, is worth at most $6,856 (4*$1,714) over four years, and is substantially less valuable than other well-known merit aid scholarships such as the Georgia HOPE and CalGrant awards (Dynarski 2008, Kane 2007). By all of these measures, the Adams Scholarship represents a relatively small amount of financial aid.

Finally, those eligible for the scholarship can use it for a maximum of eight fall and spring semesters only if they graduate from a Massachusetts public high school, are accepted at a Massachusetts public college or university, and enroll at that institution full-time by the fall following their high school graduation.[11] The student must also complete the Free Application for Federal Student Aid (FAFSA) and send the Adams Scholarship award letter to the financial aid or bursars office at the institution he or she plans to attend.[12] To continue receiving the Adams Scholarship, a student must continue his or her full-time enrollment at a Massachusetts public college or university, must maintain a cumulative college GPA of at least 3.0, and must complete the FAFSA annually.

---

[10]We spoke to financial aid officers at all of the U. Mass. campuses about their current policies, which they all believed have been in place since the inception of the Adams Scholarship. All four ask students to send their notification letters as soon as possible in the admissions process, as the financial aid offices do not have their own list of winners. U. Mass. Amherst said there was little scope for crowdout because most students send their letters after receiving financial aid offers, though students who send the letters early may be offered grant money in place of a tuition waiver. U. Mass. Lowell said that scholarship status was used in determining financial aid offers and that late notification of scholarship eligibility results in a recalculation of the aid offer. U. Mass. Boston and Dartmouth also said that scholarship status was used in determining financial aid offers but claimed that scholarship winners who would otherwise have qualified for tuition waivers would instead receive other funding.

[11]The most recent cohorts are allowed to use the scholarship within six years of graduating high school, but such cohorts are not included in our analysis.

[12]Scholarship users must also be a U.S. citizen or permanent resident of the U.S. and must have been a permanent legal resident of Massachusetts for at least one year prior to entering college as a freshman.

3.3 Data, Descriptive Statistics and College Quality

3.3.1 Data and Descriptive Statistics

The Massachusetts Department of Elementary and Secondary Education (DESE) provided the data, which include demographic information, test scores, and Adams Scholarship status for all Massachusetts public high school students expected to graduate from high school from 2004-2011. We use first time 10th grade MCAS test scores. In both math and ELA, we observe scaled scores that determine scholarship eligibility, as well as the raw scores on which those scaled scores are based. We use two main analysis samples, high school graduates from the classes of 2005-06, for whom we observe six-year college graduation rates, and high school graduates from the classes of 2005-08, for whom we observe four-year college graduation rates.[13]

College outcomes come from DESE's merge of its data on high school graduates with the National Student Clearinghouse (NSC) database, which covers 94% of undergraduates in Massachusetts.[14] We observe for each high school graduate every detailed college enrollment spell through 2012 and graduation if it occurs.[15] We add to this additional characteristics such as college costs and quality measures from the U.S. Department of Education's Integrated Postsecondary Education Data System (IPEDS) and the 2009 Barron's rankings of colleges. We separate colleges into Adams eligible institutions (U. Mass. campuses, state colleges and community colleges) and other institutions, such as in-state private or out-of-state colleges.

Table 3.1 shows the mean characteristics of the two analysis samples. Columns 1-3 contain the classes of 2005-06 and columns 4-6 contain the classes of 2005-08. Column 1 contains the full sample, column 2 limits the sample to students eligible for the Adams Scholarship, and column 3 limits the sample to those within a certain distance of the eligibility threshold, as will be

[13]We limit the sample to high school graduates, as only graduates were ultimately eligible for the Adams scholarship. Of those who receive the Adams scholarship letter in the fall of 12th grade, over 98% ultimately graduate from high school. We find no evidence that receipt of this letter affected high school graduation rates, so this restriction does not create selection bias.

[14]This figure comes from comparing NSC enrollment numbers to those contained in IPEDS. The remaining 6% come largely from for-profit institutions and those whose highest degrees take less than two years to complete. Such institutions tend to enroll students with relatively low academic skill, so that the overall match rate for those eligible for the Adams Scholarship is likely even higher than 94%. Coverage is slightly lower in the Northeast as a whole, with 90% of undergraduate enrollment covered by the NSC. Again, excluded colleges are mainly technical institutions.

[15]We exclude part-time enrollment spells and those less than 60 days long, though this has little effect on our results.

described below. Panel A shows that Adams eligible students are half as likely than the average high school graduate to be low income, black or Hispanic, because these characteristics are all negatively associated with the test scores determining eligibility. Panel B shows that 25% of high school graduates are eligible for the scholarship and that those eligible score about one standard deviation higher on their MCAS exams than the average high school graduate.

Panel C shows that 79% of Adams eligible students enroll full-time in a four-year college by the fall following their high school graduation, which we refer to as immediate enrollment. Of these, one third (26%) enroll in in-state, public, four-year colleges (Adams colleges). Panels D and E show that only 54% graduate from a four-year college within four years of high school graduation but that 71% have graduated by their sixth year. Statistics for the sample comprising the classes of 2005-08 look quite similar. Comparison of the graduation statistics to the enrollment statistics across college sectors in these samples suggest that Adams colleges have substantially lower graduation rates than do the in-state private and out-of-state colleges.

3.3.2 College Quality

Figure 3.1 confirms this difference between college sectors, plotting by initial enrollment sector the fraction of students graduating within a certain number of years. We generate these figures using NSC's data on four-year college enrollers from Massachusetts' high school class of 2004, prior to the existence of the Adams Scholarship. About 40% of those who enroll in U. Mass. campuses graduate within four years. The comparable figure for Massachusetts state colleges is well under 30%. For in-state private colleges and out-of-state colleges, that figure is about 60%. A large fraction of students in in-state public colleges use a fifth or even a sixth year to graduate. Even so, six (and even seven) years out of high school there exist large gaps in the graduation rates between these sectors. This evidence makes clear that Massachusetts' public four-year colleges have substantially longer times to degree completion and lower ultimate completion rates than the alternative colleges available to Massachusetts students.

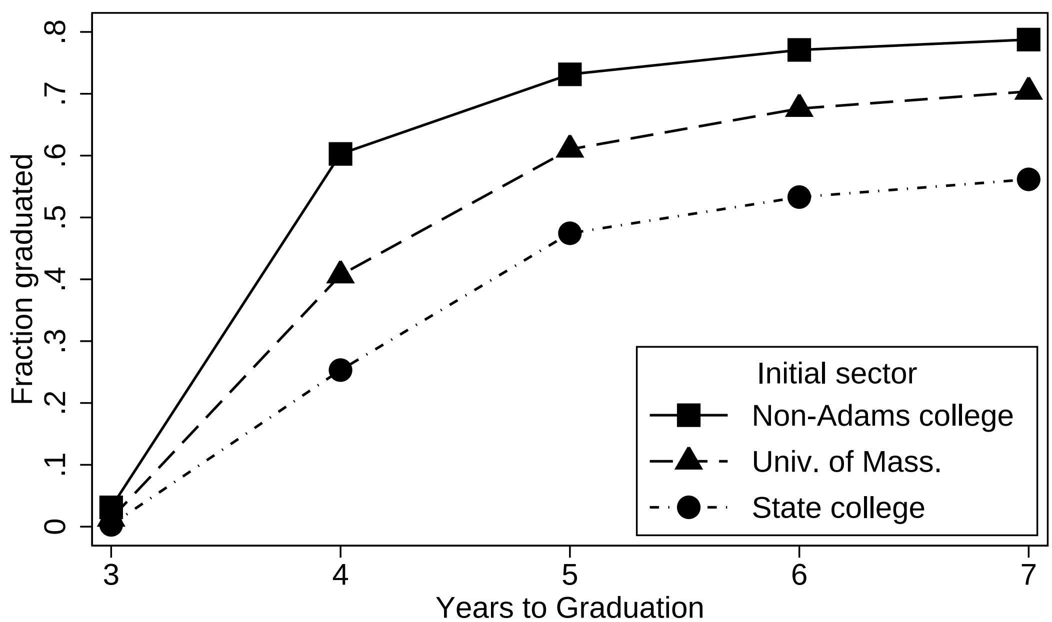
To explore why these sectors differ so dramatically in their on-time completion rates, Table 3.2 provides a more detailed description of the college market facing Massachusetts students. Quality and cost measures reported by IPEDS in the fall of 2004 are weighted by enrollment of

Table 3.1: *Summary Statistics*

	Classes of 2005-06			Classes of 2005-08		
	Full sample (1)	Adams eligibles (2)	RD sample (3)	Full sample (4)	Adams eligibles (5)	RD sample (6)
(A) Demographics						
Female	0.51	0.54	0.52	0.51	0.54	0.52
Black	0.07	0.03	0.04	0.07	0.03	0.04
Hispanic	0.07	0.03	0.03	0.07	0.03	0.04
Asian	0.05	0.07	0.05	0.05	0.07	0.05
Other race	0.00	0.00	0.00	0.01	0.01	0.01
Low income	0.17	0.09	0.10	0.18	0.10	0.11
Limited English proficient	0.03	0.01	0.01	0.02	0.01	0.01
Special education	0.12	0.01	0.03	0.13	0.01	0.03
(B) Aid eligibility						
Adams eligible	0.25	1.00	0.47	0.26	1.00	0.47
Total scaled score	491.82	527.69	515.71	494.24	527.69	516.00
Total z-score	0.19	1.07	0.78	0.18	1.02	0.74
Total z-score, Adams users	0.97	0.97	0.90	0.93	0.93	0.85
(C) Enrolled immediately						
Adams college	0.19	0.26	0.27	0.18	0.26	0.26
Non-Adams college	0.33	0.53	0.45	0.34	0.54	0.46
Four-year college	0.51	0.79	0.72	0.52	0.80	0.72
(D) Graduated within 4 years						
Adams college	0.07	0.12	0.11	0.07	0.13	0.11
Non-Adams college	0.21	0.42	0.32	0.21	0.40	0.31
Four-year college	0.28	0.54	0.43	0.29	0.52	0.42
(E) Graduated within 6 years						
Adams college	0.14	0.19	0.20	0.14	0.19	0.20
Non-Adams college	0.28	0.51	0.42	0.28	0.51	0.42
Four-year college	0.42	0.71	0.61	0.42	0.71	0.61
N	111,816	27,487	41,190	230,880	60,355	88,152

Notes: Mean values of each variable are shown by sample. Column (1) is the full sample of high school graduates from the classes of 2005-06. Column (2) restricts that sample to students eligible for the Adams Scholarship. Column (3) restricts the full sample to those within 12 points of the eligibility threshold. Columns (4)-(6) are defined similarly but for the high school classes of 2005-08. In panel (B), the last outcomes conditions on students using the Adams Scholarship to attend a four-year college. In panels (C)-(E), college outcomes all refer to four-year colleges.

Figure 3.1: *Time to Graduation by Four-Year College Sector, Class of 2004*



Notes: The above figure shows the time to graduation by initial college sector for all 2004 Massachusetts high school graduates who enroll in college immediately following high school graduation. Calculations are based on National Student Clearinghouse data.

Massachusetts students and thus represent the average student's experience of that sector. In panel A, IPEDS' measure of four-year completion rates tells a very similar story to NSC's measure, namely that U. Mass. campuses and state colleges have far lower on-time graduation rates than do non-Adams colleges.[16] Some part of this variation may be due to the academic skill of incoming students. Students enrolling in state colleges have much lower SAT scores than those enrolling in other sectors, although the U. Mass. campuses look fairly similar to non-Adams colleges in this regard. Non-Adams colleges also spend an annual average of nearly $15,000 per student on instruction, nearly twice the spending of U. Mass. campuses and more than three times the spending of state colleges. This resource gap may reduce students' access to coursework or to academic support necessary to complete such coursework and may thus help explain some of the completion rate gap. Relative to their competitors, Massachusetts' public colleges thus have substantially lower graduation rates, attract students of somewhat lower academic achievement and spend much less money on instruction.

Whether differences in graduation rates between these sectors are due to differences in incoming student achievement, resources available for instruction or other factors is beyond the scope of the paper. We follow Black and Smith (2006), who argue that because each such of these variables measures college quality with error, relationships between them and outcomes of interest will be biased toward zero. We adopt their suggestion to measure college quality by combining information from multiple variables in order to reduce such measurement error. Specifically, we construct college quality from our student-level data as the first component from a principal component analysis of each college's four-year graduation rate, SAT math 75th percentile of incoming freshmen, and instructional expenditures per student, all of which are measured by IPEDS as of 2004, prior to the Adams Scholarship. We think of the first variable as capturing the ultimate outcome of interest, the second as capturing a measure of student quality and the third as capturing a measure of available resources.[17] The first principal component from this analysis captures 64% of the variation between these three variables and nearly equally weights all three.

[16]Note that IPEDS measures the completion rate of all undergraduates in these institutions, whereas Figure 3.1 measures the completion rate only of students coming from Massachusetts public high schools.

[17]Black and Smith construct their quality measure using a slightly broader set of variables. We find that all of these quality measures are so highly correlated that it makes little difference whether we include more than three of them.

Table 3.2: *Quality and Cost by Four-Year College Sector, Class of 2004*

	Univ. of Mass. (1)	State college (2)	Non-Adams college (3)
(A) Quality			
Four-year graduation rate	0.34	0.24	0.53
SAT math 75th percentile	610	550	619
Instructional expenditures	8,224	4,342	14,510
College quality	-0.32	-0.94	0.29
(B) Costs			
Tuition	1,438	850	19,588
Required fees	6,164	3,741	666
Additional expenses	7,004	6,635	8,614
Total cost	14,606	11,224	28,867
Grant aid	6,649	5,711	14,142
Net price	7,957	5,513	14,725
Loans	3,710	2,592	4,162
N	4,828	3,488	16,881

Notes: Mean values of each variable are shown by sector for the first college of 2004 high school graduates who enroll on time in a four-year college. Quality and cost data are measured by IPEDS in the fall of 2004, with costs measured in 2004 dollars.

We standardize this quality measure to have mean zero and standard deviation one.

The final row of panel A shows that, by this measure of college quality, U. Mass. campuses and state colleges are 0.32 and 0.94 standard deviations lower than the average quality college attended by Massachusetts high school graduates. Non-Adams colleges are 0.29 standard deviations higher in quality. It is important to note here that this measure of quality is not necessarily a measure of how effectively the various college sectors are using their available resources. Though the Adams colleges have lower graduation rates and instructional expenditures, these facts may be explained in part by the fact that those colleges have much less funding per student. Panel B shows that the total cost of U. Mass. campuses and state colleges, including fees, room, board and books, are \$15,000 and \$11,000 respectively. This is about half of the \$29,000 sticker cost of their competitors.[18] When grant aid is taken into account, U. Mass. campuses charge their students an average of \$8,000 a year, relative to the \$15,000 charged by their competitors. Students, particularly those facing credit constraints, may thus make a seemingly rational decision to forgo college quality in order to attend a lower-cost public option.[19]

Table C.1 provides specific examples of four-year colleges commonly attended by the Massachusetts high school class of 2004. In 2004, U. Mass. Amherst, the college most commonly attended by Adams Scholarship recipients, had a four-year graduation rate of 43% and almost perfectly average overall quality. The other Adams colleges had substantially lower graduation rates and overall quality. Non-Adams colleges similar in graduation rate and quality to U. Mass. Amherst include Johnson & Wales University and Merrimack College in the private sector and the University of Connecticut, the University of Vermont and the University of New Hampshire in the out-of-state public sector. Elite private colleges which also enroll relatively large numbers of Massachusetts students, such as Boston University, Tufts University and Harvard University, have four-year graduation rates 50-100% higher than U. Mass. Amherst, perhaps because they attract more academically skilled students or because they spend three or more times the amount of money on student instruction.

[18]In-state community colleges, at which the scholarship could also be used, are essentially open admissions campuses. In fall 2004, they charged on average \$831 in tuition, \$2,073 in fees, and \$5,797 in other expenses, so that their sticker and net prices were roughly two-thirds those of state colleges.

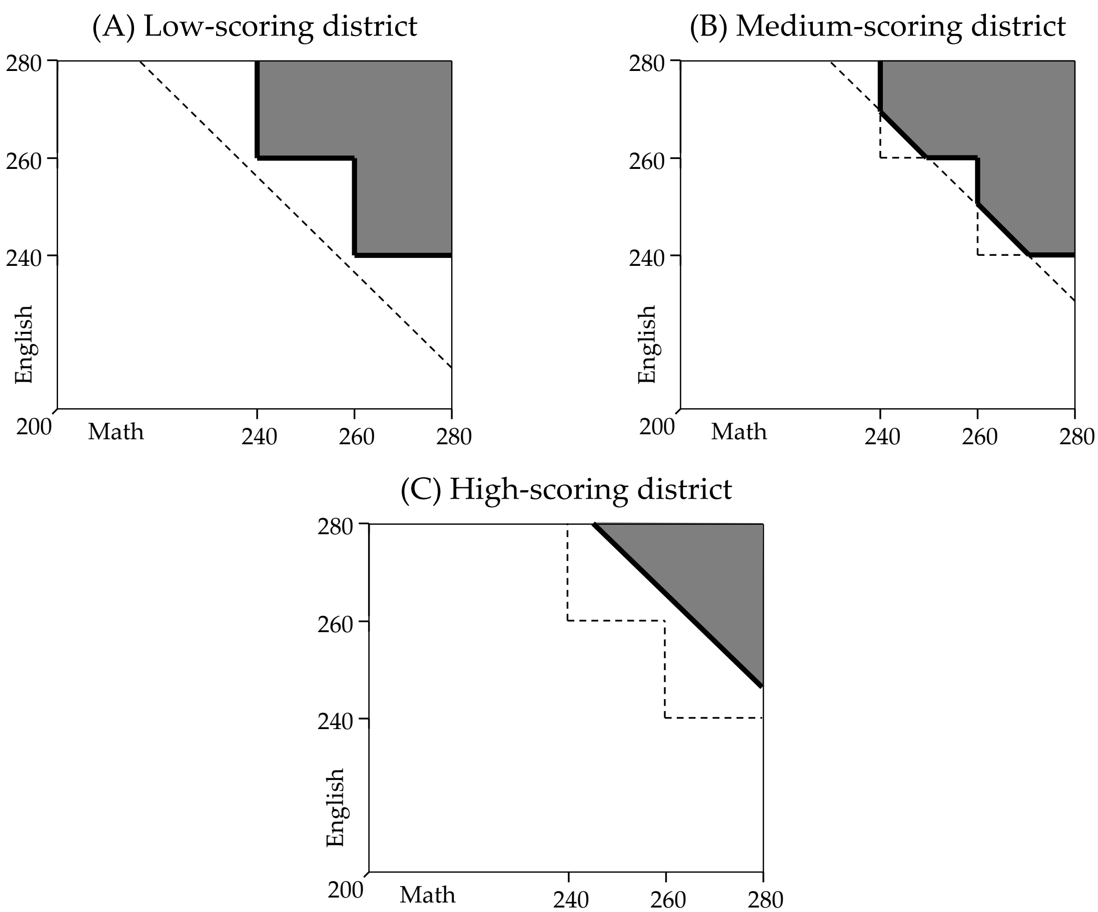
[19]In Section 3.6, we describe the average cost trade-off calculations and show that despite initial savings, there are earnings losses even larger than the savings.

3.4 Empirical Strategy

We now turn toward estimating the causal impact of the Adams Scholarship on students' college outcomes. Comparing outcomes of those eligible and ineligible for the Adams Scholarship would confound the impact of the scholarship with the fact that eligible students have higher academic skill than ineligible ones. We eliminate this source of omitted variable bias by using a regression discontinuity design that compares students just above and below the eligibility thresholds. Students just above and just below these thresholds should be similar to each other except for receipt of the scholarship. Though the scholarship may incentivize students to raise their test scores and qualify for the aid, there is little scope for manipulation of test scores around eligibility thresholds for three reasons. First, the earliest cohorts of students took their MCAS exams prior to the announcement of the Adams Scholarship. Second, at the time of test administration, the district-level 75th percentile threshold is impossible for individual students to know precisely. Third, exams are centrally scored and raw scores transformed into scaled scores via an algorithm unknown to students, their families or teachers.

Figure 3.2 provides a graphical interpretation of scholarship eligibility in three types of school districts. In each type of district, the straight line with a slope of negative one represents the cutoff that determines whether a student's total MCAS scores (math + ELA) places her in the top 25% of her school district. The W-shaped boundary defines the region in which students have scored "advanced" in one subject and "proficient" or "advanced" in the other. In low-performing districts with 25% cutoff scores of at most 500, that cutoff is so low that passing the proficient/advanced threshold is sufficient (and necessary) to win a scholarship. In medium-scoring districts with 25% cutoff scores between 502 and 518, that cutoff and proficient/advanced threshold interact in a complex way. In high-performing districts with 25% cutoff scores of at least 520, that cutoff is so high that passing it is sufficient to win. Scholarship winners are those students whose test scores thus fall in the shaded region of the graph. We note here that MCAS scores have risen dramatically since the inception of the program, as shown in Figure C.5. Because so many students pass the proficient/advanced threshold, relatively few districts in our sample are low-performing as defined by Figure 3.2. In other words, it is the top 25% boundary that is generally of the greatest importance, which can be seen by the fact that a full 25% of students qualify for the scholarship each year.

Figure 3.2: *Graphical Representation of the Eligibility Threshold*



Notes: Each panel shows a graphical representation of the set of students who are eligible for the Adams Scholarship in various types of school districts, based on scaled MCAS scores. Panel A shows a district whose 75th percentile total score is less than 500. Panel B shows a district whose 75th percentile is between 500 and 520. Panel C shows a district whose 75th percentile score is above 520.

There are many strategies for dealing with multidimensional regression discontinuities, as discussed by Reardon and Robinson (2012). Examples of such situations in the economics of education include Papay et al. (2010, 2011a,b). We collapse the discontinuity into a single dimension by defining for each student the distance of her math score from the minimum math score that defines eligibility, given her school district and ELA score. In Figure 3.2, this can be thought of as the horizontal distance between the point defined by each student's pair of test scores and the dark line defining the eligibility threshold in her school district.[20] We use raw scores rather than scaled scores in defining the running variable for two reasons. First, the raw scores are a finer measure of skill than the scaled score bins into which they are collapsed. Second, we observed extreme bunching in values of the scaled scores, particularly around the values that define the proficient and advanced thresholds. This bunching is driven entirely by the way that Massachusetts assigns groups of raw scores into scaled score bins, as the raw scores themselves have the extremely smooth distributions seen in Figures C.6 and C.7.[21]

As a result, the density of the running variable shown in Figure 3.3 looks largely smooth, suggesting little scope for endogenous manipulation that would violate the assumptions underlying the regression discontinuity design (McCrary, 2008). We do, however, see a small spike at zero itself, which is driven by the fact that a district's 75% threshold is mechanically more likely to fall on test scores that are more common in that district. Figure C.8 is consistent with this fact, showing that no such spike occurs in the low-performing districts for which only the proficient/advanced threshold, and not the 75% threshold, defines the boundary.[22] Though the spike is small and not driven by endogenous manipulation of the running variable itself, we later show that our central results are robust to and even strengthened by excluding students directly on the boundary, in a so-called "doughnut hole" regression discontinuity.

To estimate the causal effect of the Adams Scholarship, we use local linear regression to estimate linear probability models of the form:

$$Y_{ijt} = \beta_0 + \beta_1 Adams_{ijt} + \beta_2 Gap_{ijt} + \beta_3 Gap_{ijt} \times Adams_{ijt} + \epsilon_{ijt}. \tag{3.1}$$

where Gap_{ijt} is the running variable described above and *Adams* is an indicator for Adams Scholarship eligibility ($Gap_{ijt} \geq 0$).[23] The causal effect of winning the Adams Scholarship on

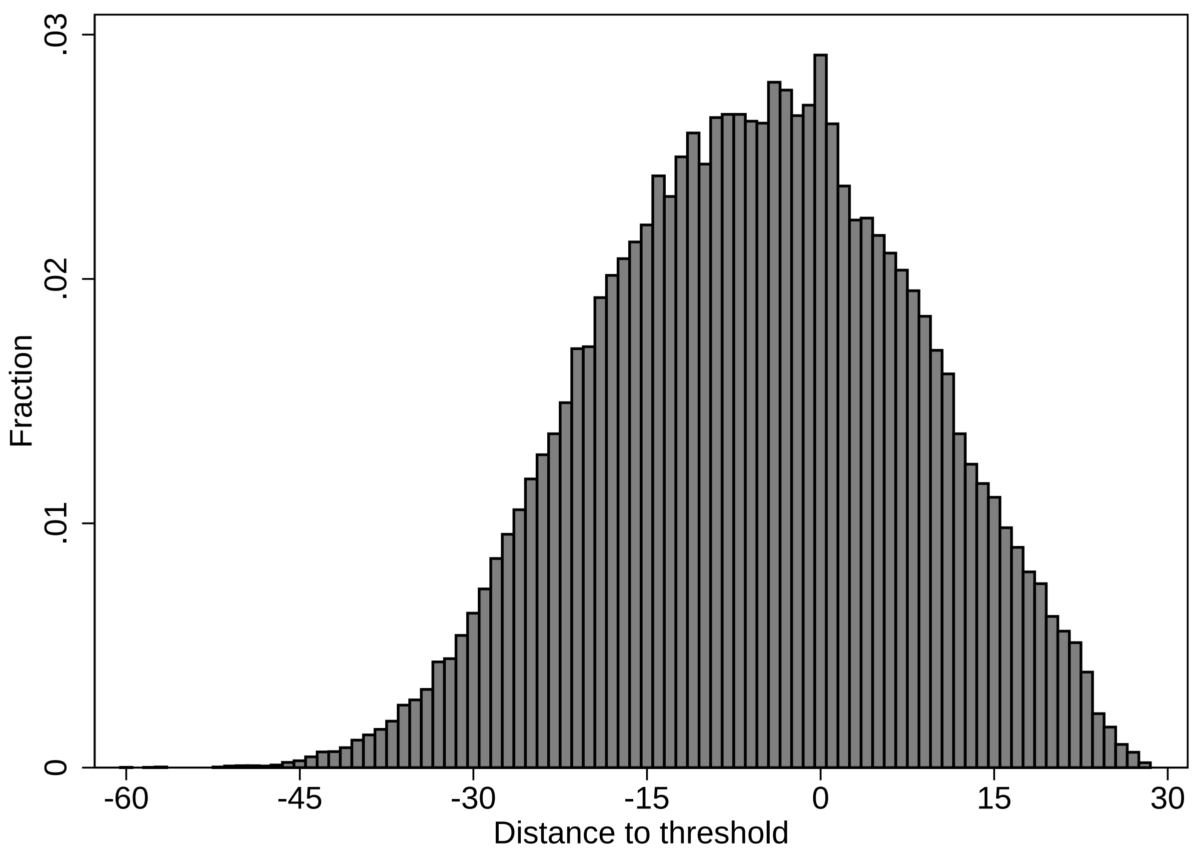
[20] Our results are robust to defining the running variable as the vertical distance, the distance of each student's ELA score from the minimum ELA score that defines eligibility, given her school district and math score.

[21] Goodman (2008) characterized each student by the minimum of her scaled score distance from the proficient/advanced and top 25% thresholds. Distance to the top 25% threshold is not an easily defined quantity when raw scores are used because the straight line boundary observed in Figure 3.2 becomes quite jagged. We therefore prefer the running variable described in the text above. Estimates using the running variable as defined in Goodman (2008) are, nonetheless, quite similar to those presented here and are available by request from the authors.

[22] Figure C.9 show very similar patterns for the 2005-08 sample.

[23] We use linear probability models here and in our later IV regressions rather than limited dependent variable

Figure 3.3: *Density of Running Variable, Classes of 2005-06*



Notes: The above figure shows the distribution of the running variable for the high school classes of 2005-06. The running variable is defined as the distance of a given student's raw math score from the raw math score defining her school district's threshold, given her raw ELA score.

an outcome, Y_{ijt}, should be estimated by β_1 if the usual assumptions underlying the validity of the regression discontinuity design are not violated. Assuming that treatment effects are homogeneous along different parts of the eligibility threshold, this coefficient measures a local average treatment effect for students near the threshold, weighted by the probability of a given student being near the threshold itself (Reardon and Robinson, 2012).

Our preferred implementation uses local linear regression with an triangular kernel that weights points near the threshold more heavily than those far from the threshold. We compute optimal bandwidths following the procedure developed by Imbens and Kalyanaraman (2012), which trades off precision for bias generated by deviations from linearity away from the threshold. Across nearly all of our outcomes and samples, the optimal bandwidth generated by this procedure falls somewhere between 10 and 15 raw score points. For simplicity and ease of defining a single sample across outcomes, we choose as our default specification a bandwidth of 12. We then show that our results are quite robust to a wider set of bandwidths, to inclusion of demographic controls, to inclusion of school district by cohort fixed effects, and to use of parametric specifications, including polynomials of various degrees. We cluster standard errors by 12th grade school district in all specifications in order to account for within district correlations in the error term ϵ_{ijt}.

As further reassurance of the validity of the discontinuity design employed here, Table 3.3 tests whether observed covariates vary discontinuously at the eligibility threshold. The first eight columns test the basic covariates, including gender, race, low income, limited English proficiency and special education status. With the exception of marginally significant but small differences in the probability of being black or "other" race for the 2005-06 sample, none of those covariates shows a statistically significant discontinuity in either the 2005-06 or the 2005-08 sample. The estimates are precise enough to rule out economically significant discontinuities as well. To test whether these covariates are jointly discontinuous, we generate in columns 9 and 10 predicted math and ELA z-scores by regressing scores from the class of 2004 on the demographic controls listed in the previous eight columns. We then use the resulting regression estimates to predict scores for students in subsequent classes. The estimates in columns 9 and 10 suggest no

models for the reasons discussed by Angrist (2001). In particular, we are interested in directly interpretable causal effects and not on structural parameters generated by non-linear models. We also note that estimates generated by probit and logit models turn out to be extremely similar to those generated by the linear probability model above.

discontinuity in predicted test scores and the estimates are precise enough to rule out differences around the eligibility threshold of more than 0.02 standard deviations in academic skill. Figure 3.4 shows graphically the average predicted scores of students in each bin defined by distance from the eligibility threshold, confirming the lack of any clear difference in academic skill between students just above and just below the threshold in the 2005-06 sample.[24]

3.5 Results

3.5.1 Enrollment and Graduation Rates

To visualize the enrollment impacts of the Adams Scholarship, we plot in Figure 3.5 the proportion of 2005-06 graduates for each value of Gap who enroll in four-year colleges immediately following high school graduation.[25] There is clear visual evidence that students at the eligibility threshold are substantially more likely to enroll in an Adams (i.e., in-state public) college than students just below the threshold. Such students are, however, similarly less likely to enroll in a non-Adams (i.e., in-state private or out-of-state) college, the net result of which is little apparent difference in overall college enrollment rates between these two groups of students.

The first row of Table 3.4 reports estimates of these differences in the 2005-06 sample. Scholarship eligibility induced 6.9 percent of students at the threshold to enroll in Adams colleges, a more than one-fourth increase over the 23.8 percent enrollment rate of students just below the threshold.[26] More than six-sevenths of these marginal students, or 6.0 percent, would have attended other four-year colleges if not for the scholarship. The net result is a statistically insignificant 0.9 percentage point increase in the fraction of students enrolling in any four-year college. Many of these marginal students switched their enrollment from out-of-state colleges, leading to a 4.8 percentage point increase in the fraction of students enrolling in-state four-year colleges. The Adams Scholarship therefore did induce a substantial number of students to enroll in the public sector and succeeded in keeping some students in state who otherwise would have

[24]The 2005-08 sample looks quite similar, as seen in Figure C.10.

[25]Immediate enrollment was a requirement of the scholarship.

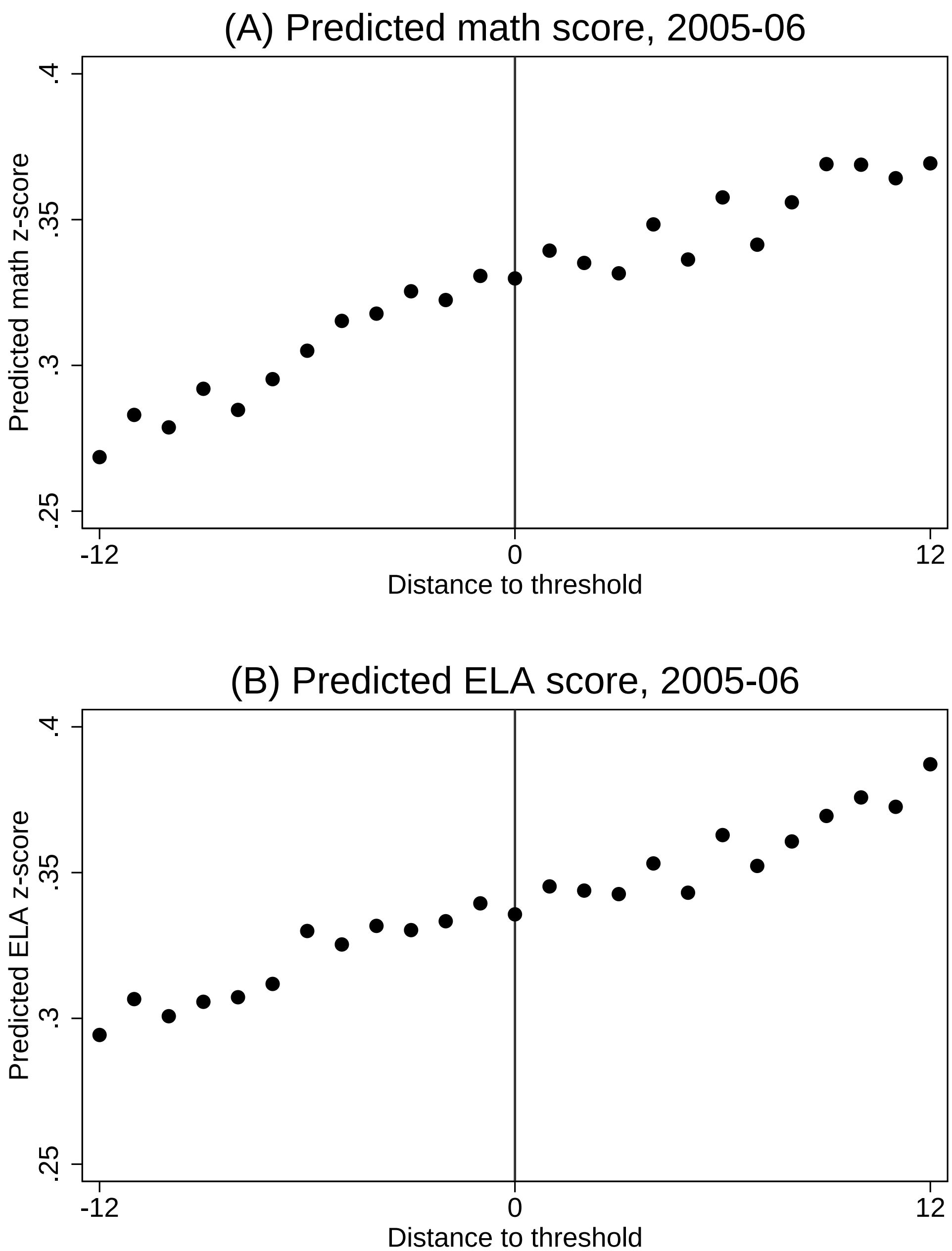
[26]Table C.2 shows that nearly half of these marginal students enrolled in U. Mass. Amherst, the flagship campus, and another third enroll in the various state colleges.

Table 3.3: *Covariate Balance*

	Female (1)	Black (2)	Hispanic (3)	Asian (4)	Other race (5)	Low income (6)	Ltd. Eng. prof. (7)	Special ed. (8)	Predicted math score (9)	Predicted ELA score (10)
(A) Classes of 2005-06										
Adams eligible	-0.005	0.006*	-0.000	-0.005	-0.002*	0.003	-0.000	0.003	-0.007	-0.006
	(0.010)	(0.004)	(0.004)	(0.005)	(0.001)	(0.006)	(0.001)	(0.003)	(0.005)	(0.005)
$\bar{Y}$	0.508	0.038	0.033	0.052	0.005	0.103	0.005	0.018	0.331	0.339
N	41,190	41,190	41,190	41,190	41,190	41,190	41,190	41,190	41,190	41,190
(B) Classes of 2005-08										
Adams eligible	0.008	0.002	-0.001	0.000	-0.000	0.000	0.000	-0.000	-0.002	0.000
	(0.007)	(0.002)	(0.003)	(0.003)	(0.001)	(0.004)	(0.001)	(0.002)	(0.003)	(0.003)
$\bar{Y}$	0.498	0.041	0.040	0.051	0.008	0.118	0.004	0.026	0.313	0.322
N	88,152	88,152	88,152	88,152	88,152	88,152	88,152	88,152	88,152	88,152

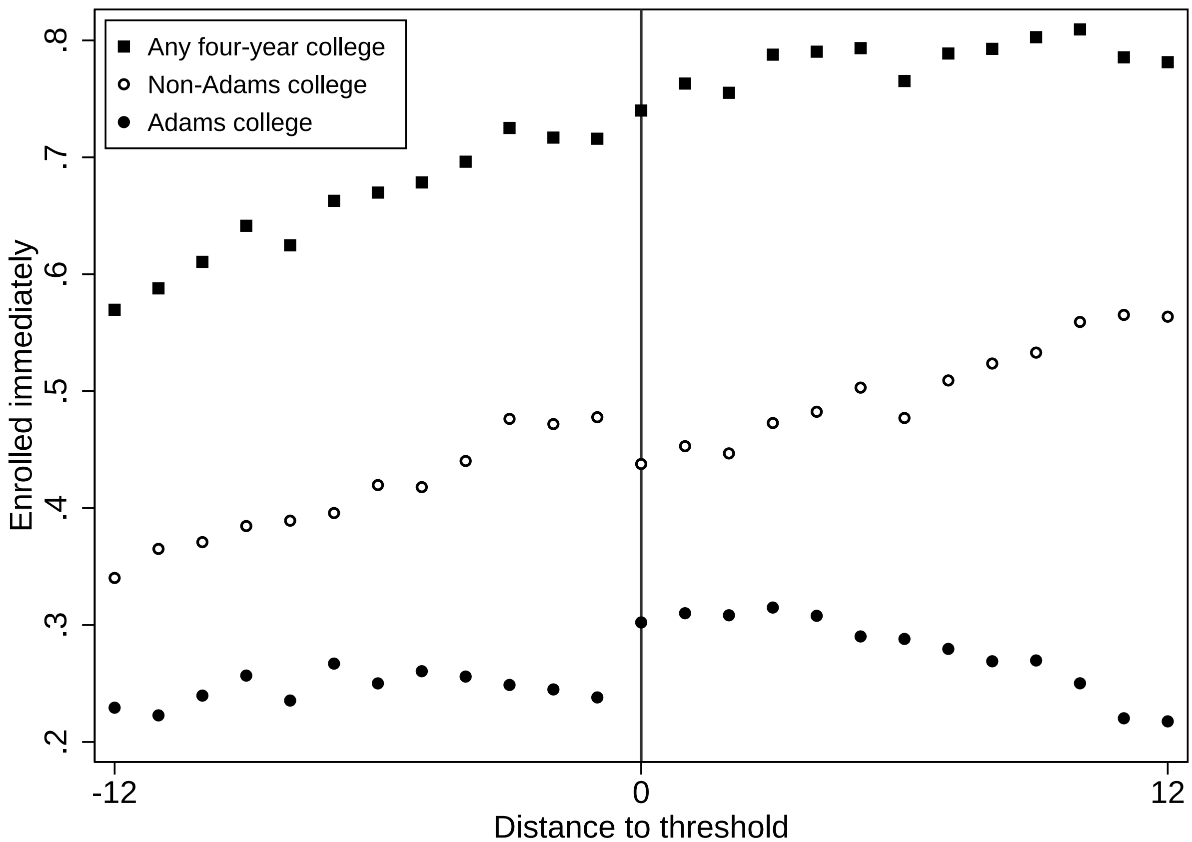
Notes: Heteroskedasticity robust standard errors clustered by 12th grade school district are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). Demographic controls listed at the top of each column are used as outcomes. Predicted math and ELA scores used as outcomes in the last two columns are generated by regressing standardized scores from the class of 2004 on the demographic controls listed in the previous eight columns. We then use the resulting estimates to predict scores for students in subsequent classes. Each coefficient on aid eligibility is generated by local linear regression with a triangular kernel of bandwidth 12. In panel (A), the sample consists of the high school classes of 2005-06. In panel (B), the sample consists of the high school classes of 2005-08. Listed below each coefficient is the mean of the outcome for students just below the eligibility threshold.

Figure 3.4: *Smoothness of Covariates, Classes of 2005-06*



Notes: Each panel shows the mean predicted math and ELA score by each value of the running variable, for the high school classes of 2005-06. Predicted scores are generated by regressing math and ELA scores on demographic characteristics for the class of 2004. The resulting coefficients are then used to generate predictions for subsequent classes.

Figure 3.5: *Enrollment at Four-Year Colleges, Classes of 2005-06*



Notes: The above figure shows the fraction of students enrolling in four-year colleges immediately following high school graduation by each value of the running variable, for the high school classes of 2005-06. Adams colleges are Massachusetts public four-year colleges where the Adams Scholarship tuition waiver may be used. Non-Adams colleges are all other four-year colleges, both in-state and out of state. Calculations are based on National Student Clearinghouse data.

left.

The scholarship also induces a statistically insignificant 0.8 percentage point increase in the fraction of students who enrolled in two-year community colleges. That, combined with the slight rise in four-year college enrollment rates, implies that the scholarship raised overall immediate college enrollment rates by 1.7 percentage points. In the second row, we define as the outcome enrollment within two years of high school graduation, rather than immediately following graduation. The estimates in columns 1 and 3 fall by 0.6 and 0.9 percentage points respectively, suggesting that a small number of marginal students induced to enroll immediately in Adams colleges because of the scholarship would have enrolled within the next two years in the absence of the scholarship. The estimates in column 3 suggest that the scholarship may have accelerated enrollment in four-year colleges for a small number of students but did not induce enrollment in four-year colleges for any students who would not have enrolled within two years. Interestingly, the two-year college effect is unchanged by the shift in definition from immediate enrollment to enrollment within two years. This suggests that scholarship eligibility may have induced a small number of students to enroll in community colleges who would not otherwise have enrolled within two years.

Turning from enrollment to graduation, we plot in Figure 3.6 the proportion of students for each value of *Gap* who graduate from four-year colleges within six years of high school graduation. Students just above the eligibility threshold are more likely to have graduated from Adams colleges than those just below the threshold, an unsurprising result given that the former are much more likely to enroll in that sector than the latter. Scholarship eligibility also lowers graduation rates from non-Adams colleges, for the same reason that eligibility reduces initial enrollment in that sector. More surprising is that the decrease in graduation rates from non-Adams colleges is larger in magnitude than the increase in graduation rates from Adams colleges. The net result is that scholarship eligibility lowers overall graduation rates, as shown by the top line in Figure 3.6, where points to the right of the eligibility threshold are generally lower than would be predicted if extrapolating from points to the left of the threshold.[27]

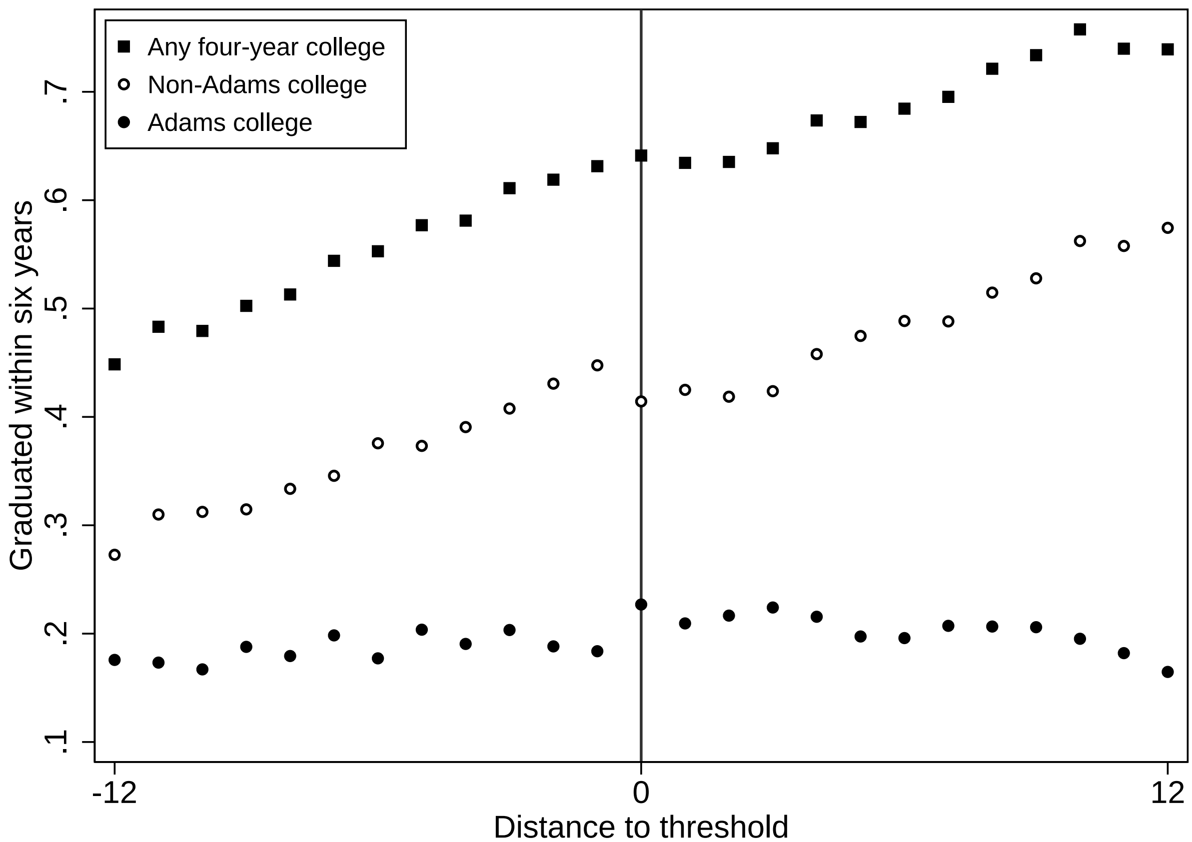
[27]One anomaly is that students exactly on the threshold ($GAP = 0$) have higher graduation rates than students just to the left of the threshold ($GAP = -1$). We believe this is either an artifact of noisy data or driven by the slight bunching at zero described earlier in the text. Regression estimates that we discuss below, which show a clear discontinuity in the overall graduation rate, become even larger in a "doughnut hole" regression that excludes students

Table 3.4: *Impact of Aid Eligibility on Enrollment and Graduation, Classes of 2005-06*

	Adams college, four-year (1)	Non-Adams college, four-year (2)	Any four-year college (3)	In-state, four-year (4)	Any two-year college (5)	Any college (6)
Enrolled immediately	0.069***	-0.060***	0.009	0.048***	0.008	0.017**
	(0.010)	(0.010)	(0.008)	(0.010)	(0.005)	(0.008)
$\bar{Y}$	0.238	0.478	0.716	0.409	0.068	0.784
Enrolled within 2 years	0.063***	-0.062***	-0.000	0.031***	0.007	0.007
	(0.010)	(0.011)	(0.007)	(0.010)	(0.005)	(0.006)
$\bar{Y}$	0.292	0.547	0.796	0.502	0.073	0.870
On campus, year 4	0.030***	-0.053***	-0.023***	0.003	0.004	-0.020**
	(0.009)	(0.010)	(0.009)	(0.009)	(0.004)	(0.009)
$\bar{Y}$	0.230	0.456	0.685	0.425	0.044	0.730
Graduated within 4 years	0.020***	-0.037***	-0.017*	0.006	-0.002	-0.019*
	(0.006)	(0.010)	(0.010)	(0.008)	(0.003)	(0.010)
$\bar{Y}$	0.095	0.338	0.433	0.216	0.029	0.462
Graduated within 6 years	0.029***	-0.053***	-0.025***	0.003	-0.000	-0.025***
	(0.008)	(0.010)	(0.009)	(0.009)	(0.004)	(0.009)
$\bar{Y}$	0.184	0.448	0.631	0.367	0.031	0.662
N	41,190	41,190	41,190	41,190	41,190	41,190

Notes: Heteroskedasticity robust standard errors clustered by 12th grade school district are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient on aid eligibility is generated by local linear regression using a triangular kernel of bandwidth 12. The sample consists of the high school classes of 2005-06. In the first row, the outcome is defined as enrollment by the fall following high school graduation. In the second row, the outcome is defined as enrollment within two years of the fall following high school graduation. In the the third row, the outcome is defined as being enrolled in college in the spring of the fourth year after high school graduation. In the fourth and fifth rows, the outcome is defined as college graduation within four and six years of high school graduation, respectively. Listed below each coefficient is the mean of the outcome for students just below the eligibility threshold.

Figure 3.6: *Graduation from Four-Year Colleges, Classes of 2005-06*



Notes: The above figure shows the fraction of students graduating from four-year colleges within six years of high school graduation by each value of the running variable, for the high school classes of 2005-06. Adams colleges are Massachusetts public four-year colleges where the Adams Scholarship tuition waiver may be used. Non-Adams colleges are all other four-year colleges, both in-state and out of state. Calculations are based on National Student Clearinghouse data.

The third through fifth rows of Table 3.4 confirm this decrease in graduation rates. The third row uses as an outcome an indicator for the student being enrolled in a given college sector as of the spring of the fourth year after her high school graduation, which we interpret as a measure of persistence. The fourth and fifth rows use as outcomes indicators for whether a student has graduated from a given college sector within four or six years. The three rows tell a consistent story. Though scholarship eligibility increased enrollment in Adams colleges by nearly seven percentage points, it increased persistence and six-year graduation rates by only three percentage points, suggesting that the majority of marginal students did not successfully graduate from that sector. Scholarship eligibility reduced persistence and graduation rates in the private sector by over five percentage points. The net result is that scholarship eligibility reduced the probability of earning a four-year college degree within six years by 2.5 percentage points. That the persistence and four-graduation rate measures show similar declines suggests this is not merely a matter of delaying graduation but instead is driven by a subset of students who have dropped out of the four-year college sector entirely.

We note three other important findings. First, although scholarship eligibility increased the number of students enrolling in state, it had no ultimate effect on the probability of earning a degree in state. Second, none of the increased enrollment in community colleges translated into increased completion of two-year college degrees, even six years out of high school. Third, as a result, scholarship eligibility lowered by 2.5 percentage points the probability that a student had any college degree six years after high school graduation.

Table 3.5 explores these enrollment and graduation impacts over time, with the first four columns analyzing each high school class separately, the fifth pooling the classes of 2005-08, and the sixth showing enrollment effects for the classes of 2009-11, the most recent for which data are available. Panel A shows that scholarship eligibility increased enrollment in four-year Adams colleges for all graduating high school classes. There is, however, a gradual monotonic decrease in the impact of scholarship over time, with the effect in 2005 three times that of the effect over 2009-11. This gradually shrinking effect size may be driven by the fact that rapidly rising fees have shrunk the proportion of college costs covered by the scholarship. Also worth noting is that

on the threshold. For evidence of this, see Table C.3. Though not show here, difference-in-difference estimates that use the 2004 cohort as a pre-policy control group show similarly negative impacts on graduation rates, confirming that students to the right of the threshold are graduating at lower rates than would otherwise be predicted.

much of increase in overall four-year college enrollment is driven by the first treated class, with subsequent classes showing smaller and insignificant impacts on this margin.

Table 3.5: *Enrollment, Persistence and Graduation by High School Class*

	2005 (1)	2006 (2)	2007 (3)	2008 (4)	2005-08 (5)	2009-11 (6)
(A) Enrolled						
Immediately, Adams college	0.073***	0.067***	0.059***	0.046***	0.060***	0.020***
	(0.016)	(0.012)	(0.013)	(0.012)	(0.007)	(0.007)
$\bar{Y}$	0.234	0.242	0.223	0.241	0.235	0.234
Immediately, four-year college	0.029**	-0.005	0.013	0.015	0.012**	-0.006
	(0.014)	(0.011)	(0.012)	(0.011)	(0.006)	(0.007)
$\bar{Y}$	0.692	0.733	0.730	0.716	0.719	0.713
Within 2 years, four-year college	0.017	-0.014	-0.000	0.004	0.001	-0.009
	(0.012)	(0.009)	(0.012)	(0.010)	(0.005)	(0.006)
$\bar{Y}$	0.770	0.816	0.787	0.775	0.788	0.762
(B) Graduated						
On campus, year 4	-0.003	-0.039***	-0.018	-0.008	-0.017***	
	(0.016)	(0.011)	(0.014)	(0.011)	(0.006)	
$\bar{Y}$	0.679	0.690	0.631	0.620	0.653	
Within 4 years	-0.017	-0.018	-0.008	-0.002	-0.010	
	(0.016)	(0.013)	(0.014)	(0.013)	(0.007)	
$\bar{Y}$	0.418	0.444	0.411	0.420	0.424	
Within 5 years	-0.018	-0.034***	-0.016		-0.023***	
	(0.016)	(0.013)	(0.014)		(0.008)	
$\bar{Y}$	0.570	0.595	0.562		0.576	
Within 6 years	-0.024	-0.026**			-0.025***	
	(0.015)	(0.012)			(0.009)	
$\bar{Y}$	0.630	0.632			0.631	
N	18,270	22,920	21,808	25,154	88,152	74,139

Notes: Heteroskedasticity robust standard errors clustered by 12th grade school district are in parentheses (* p<.10 ** p<.05 *** p<.01). Each coefficient on aid eligibility is generated by local linear regression using a triangular kernel of bandwidth 12. Each column consists of a different high school class or set of classes. In panel (A), the outcomes are defined as enrollment immediately or within two years of high school graduation. In panel (B), the outcomes are defined as being on campus or graduating from any four-year college, regardless of initial enrollment choice. Listed below each coefficient is the mean of the outcome for students just below the eligibility threshold.

Panel B estimates the impact of scholarship eligibility on persistence and graduation after four, five and six years. Three findings are worth noting. First, that the magnitude of the persistence and various graduation rates do not vary much within classes implies that the negative impact of scholarship eligibility on graduation rates is driven largely by dropout rather than delay. Second, that the negative graduation effect is not driven solely by the first high school class makes much less likely the possibility that the effect was generated by confusion about the meaning of "free tuition" in the scholarship letter. If such language was deceiving students into making uninformed decisions, we would expect such negative graduation effects to diminish across classes as information about the true value of the scholarship spread. There is no clear evidence of such a pattern. Third, estimated impacts on enrollment and persistence rates generated by the full 2005-08sample are similar to those generated by the 2005-06 sample. Figures C.11 and C.12 confirm this, replicating Figures 3.5 and 3.6 for the larger sample. The two sets of figures look quite similar. As a whole, this evidence suggests a fairly stable impact of the scholarship on enrollment, persistence and graduation.

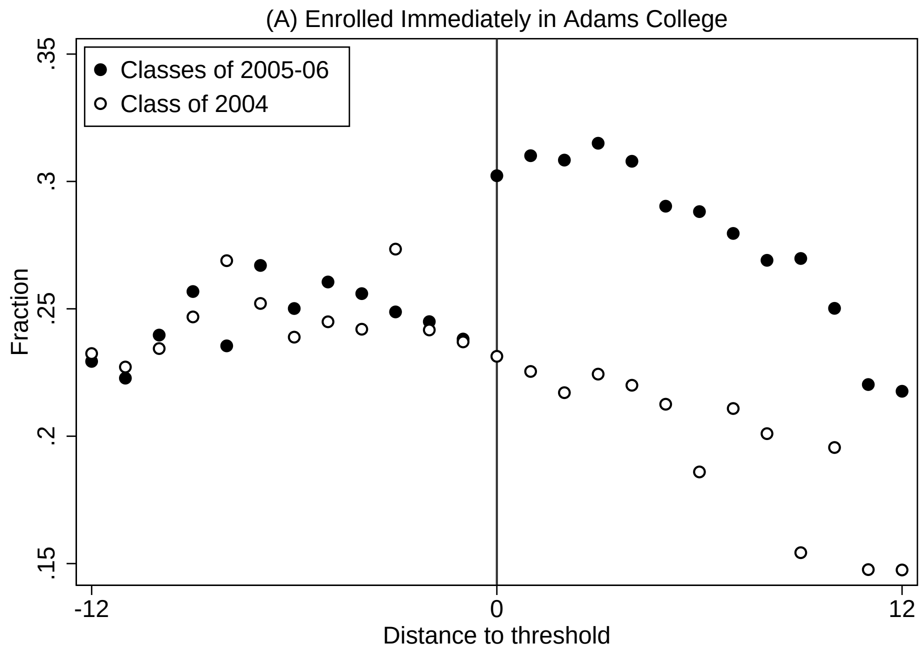
Table C.3 tests the robustness of our central results to a variety of alternative specifications. Panel A replicates our default local linear regression specification for a variety of bandwidths, beginning with the Imbens-Kalyaramanan optimal bandwidth. In the 2005-06 sample, that optimal bandwidth is about 14 for enrollment outcomes and about 10 for persistence and graduation outcomes, hence our choice of 12 as the default bandwidth in earlier tables. The magnitude and statistical significant of these estimates are generally quite robust to these changes in the bandwidth.

Panel B replicates our default specification, using a bandwidth of 12, with three variations. First, we include school district by high school class fixed effects to account for the fact that the eligibility threshold differs by district and class. This has very little impact on the estimates. Second, we include demographic controls, which also change the estimates very little. Third, we run a doughnut hole regression in which we exclude students exactly on the boundary, because of the small amount of bunching observed in the running variable. This actually increases the magnitude of our enrollment and graduation estimates by roughly a percentage point, suggesting that the mild bunching was, if anything, causing us to underestimate the impacts of the scholarship. Finally, panel C fits quadratic, cubic and quartic polynomials on either side of the threshold, using the entire sample and a rectangular kernel. This yields similar estimates to those generated by the local linear regression used as our default specification.

As a final piece of evidence, Figure 3.7 exploits as a placebo test data from the high school class of 2004, the one cohort in our data that graduated prior to the scholarship's existence. In panel A, we see no evidence of a discontinuity in Adams college enrollment for the class of 2004, compared to the previously observed clear discontinuity for the classes of 2005-06. Similarly, panel B shows that students below the threshold have similar six-year graduation rates across the three classes, whereas students above the threshold in 2005-06 have lower rates than such students in 2004. That the discontinuities in enrollment and graduation appear only in the years when the scholarship existed strengthens the case that it is due to the policy itself and not other unobserved factors.

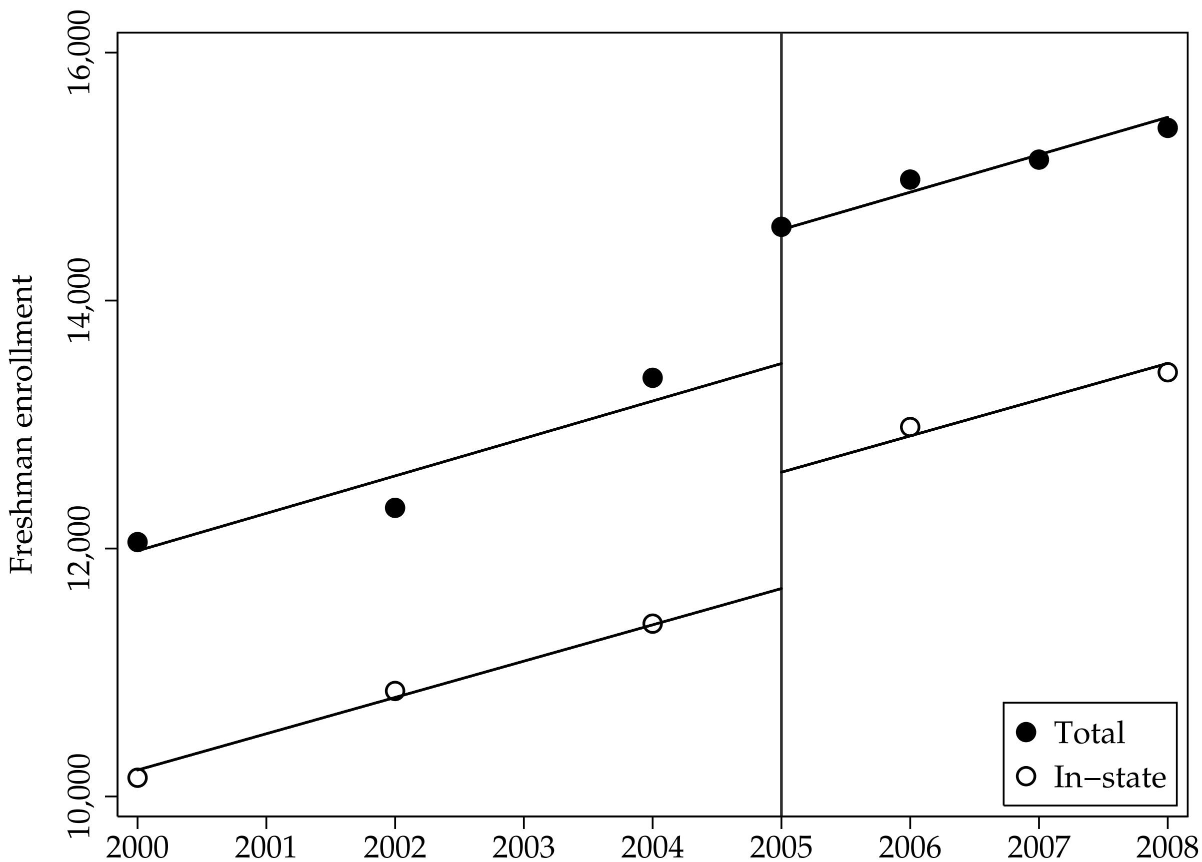
Panel A also highlights that the magnitude of the enrollment impact is large even for students somewhat far from the threshold. Our regression discontinuity estimates, as well as those based on difference-in-difference calculations following Figure 3.7, suggest that the Adams Scholarship induced about 1,000 additional students to enroll in in-state public colleges. IPEDS data reported by Massachusetts' public colleges themselves confirms this. Figure 3.8 plots the reported freshman enrollment across all Massachusetts public four-year colleges, both for all students and for those from Massachusetts. There is a clear trend break in 2005, when the Adams Scholarship begins, due entirely to increased numbers of Massachusetts freshman and of magnitude nearly identical to our estimate. This implies that the additional students induced into in-state public colleges did not crowd out other students, instead simply adding to each campus at most a few hundred students who would not otherwise have enrolled there.

Figure 3.7: *Treatment vs. Pre-Treatment Classes*



Notes: Panel A shows shows the fraction of students enrolling in a Massachusetts public four-year college immediately following high school graduation, for the treated high school classes of 2005-06 and the untreated class of 2004. Panel B shows the fraction of students graduating from any four-year college within six years of high school graduation. Calculations are based on National Student Clearinghouse data.

Figure 3.8: *Freshman Enrollment in Four-Year Adams Colleges*



Notes: The above figure shows the total number of freshmen enrolled in Massachusetts public four-year colleges, as well as the number of such freshmen coming from Massachusetts. Calculations are based on IPEDS data.

In summary, the primary effect of the Adams Scholarship was to induce large numbers of students to switch into in-state public four-year colleges from other four-year colleges they otherwise would have attended, a result consistent with Goodman (2008). The scholarship did increase in-state college enrollment rates but had little impact on in-state graduation rates. Scholarship eligibility actually reduced overall graduation rates, for reasons we now turn to.

3.5.2 College Quality and Cost

The most plausible explanation for the negative impacts on graduation rates is that the scholarship induced students to attend colleges with substantially lower graduation rates than they otherwise would have. Table 3.6 explores the quality and cost tradeoffs that the Adams Scholarship induced. The top row presents reduced form estimates of the impact of scholarship eligibility on a variety of college quality and cost measures, as in Equation 3.1 above. For this analysis, we assign students to the four-year college to which they enroll immediately following high school graduation. The bottom row estimates these impacts for the marginal student using the following equations that instrument enrollment in an in-state public college with scholarship eligibility:

$$Y_{ijt} = \beta_0 + \beta_1 Adams\hat{C}ollege_{ijt} + \beta_2 Gap_{ijt} + \beta_3 Gap_{ijt} \times Adams_{ijt} + \epsilon_{ijt} \tag{3.2}$$

$$AdamsCollege_{ijt} = \alpha_0 + \alpha_1 Adams_{ijt} + \alpha_2 Gap_{ijt} + \alpha_3 Gap_{ijt} \times Adams_{ijt} + \nu_{ijt} \tag{3.3}$$

In the first column, we generate an indicator for a college being highly competitive if Barron's 2009 rankings placed that college into one of its top three categories of "most competitive," "highly competitive," and "very competitive,". None of Massachusetts' public colleges fall into these categories, which include colleges such as Boston University, Tufts University, Simmons College, and Lesley University. All of the U. Mass. campuses and nearly all of the state colleges fall into the fourth category of "competitive," which also includes private colleges such as Suffolk University and the Wentworth Institute of Technology. The fifth category of "not competitive" includes two state colleges and all community colleges. Column 1 shows that, for the classes of 2005-06, scholarship eligibility induced an estimated 3.3% of students, or 48% of those switching colleges, to forgo institutions in those highest three categories. Students did not simply switch into the public sector from private or out-of-state colleges of similar quality. Half of the students induced to switch colleges would have enrolled in more competitive alternatives in the absence of the scholarship.

Other measures of college quality, which are defined only for students immediately enrolling in four-year colleges, point to a similar pattern. In column 2, the estimates suggest that students induced by the scholarship to switch into Adams colleges would otherwise have attended colleges with four-year graduation rates nearly 17 percentage points higher. These marginal students would also have attended colleges with higher SAT math scores (by 27 points) and higher instructional spending per student (by $3,700 annually), though this last estimate is not statistically significant. Combining these three measures as described above, column 5 shows that scholarship eligibility

Table 3.6: *Impact of Aid on Initial College Quality and Cost*

	Highly comp. (1)	Four-year grad. rate (2)	SAT math (3)	Instr. spending (4)	College quality (5)	Net price (6)	Adams aid (7)
Adams eligible (RF)	-0.033***	-0.015***	-2.450	-0.332	-0.057***	-0.929***	0.094***
	(0.010)	(0.005)	(1.502)	(0.283)	(0.022)	(0.139)	(0.014)
Adams college (IV)	-0.477***	-0.166***	-27.163*	-3.684	-0.637***	-10.308***	1.353***
	(0.137)	(0.042)	(15.711)	(2.973)	(0.203)	(0.757)	(0.052)
$\bar{Y}$	0.323	0.488	621.228	13.638	0.177	15.258	0.323
St.Dev.		0.210	62.357	13.364	0.963	5.846	
N	41,190	29,639	29,639	29,639	29,639	29,635	41,190

Notes: Heteroskedasticity robust standard errors clustered by 12th grade school district are in parentheses (* $p<.10$ ** $p<.05$ *** $p<.01$). The sample consists of the high school classes of 2005-06. In the top row, each coefficient on aid eligibility is generated by local linear regression with a triangular kernel of bandwidth 12. The bottom row instruments enrollment in a four-year Adams college with aid eligibility, using the same specification. The highly competitive category includes institutions in the top three Barron's categories (most, highly, and very competitive). The remaining outcomes are defined only for students who enrolled immediately in four-year colleges. College quality is the standardized first principal component of each institution's four-year graduation rate, the 75th percentile SAT math score, and instructional expenditures per student, measured as of 2004. Adams aid is measured for the marginal student by imputing aid to all students below the threshold attending Adams colleges. All financial outcomes are measured in thousands of dollars. Listed below each coefficient is the mean and standard deviation of the outcome for students just below the eligbility threshold.

induced the marginal student to forgo more than 0.6 standard deviations in college quality.

In exchange for this drop in quality, students enrolled in public colleges where the average student's net price of attendance was $10,000 a year lower than private and out-of-state alternatives, as seen in column 6. This is the most direct measure we provide of the extent to which this in-kind subsidy reduces consumption of college education, as in Peltzman (1973). This cost difference would, however, have been available to these students even in the absence of the Adams Scholarship. The scholarship itself was worth, on average, less than $1,400 a year to such students, as seen in column 7.[28] Combining the estimates from columns 5 and 7 suggests a willingness to forgo 0.47 (0.637/1.353) standard deviations of college quality per $1,000 in annual aid. Estimates for the larger sample of 2005-08 graduates are quite similar. Below, we calculate the estimated impacts of such a tradeoff on lifetime earnings and find that these responses are hard to explain in classical human capital model. Students seem remarkably willing to forgo college quality and attend institutions with low graduation rates in exchange for relatively small amounts of financial aid.

To strengthen our case that the decrease in college quality induced by the scholarship explains the observed graduation impacts, we explore heterogeneity by a variety of characteristics in Table 3.7. Here we use the classes of 2005-08 to improve precision among relatively small subgroups. Panel A takes advantage of the fact that the academic skill level defined by eligibility threshold varied by school district due to the requirement that students be in the top 25% of their district peers. We therefore divide districts into quintiles by the fraction of 2004 graduates who attended a competitive college, as defined previously. We then fully interact our baseline specification from prior tables with indicators for being from the bottom quintile, middle three quintiles, and top quintile of school districts, and also include the direct effects of those indicators.

For students from the bottom quintile districts, who are on average lower income and less academically skilled, scholarship eligibility increases enrollment in Adams colleges by nearly eight percentage points, of which five percentage points represent students who would not otherwise have enrolled immediately in any four-year college. Eligibility does not, however, reduce such students' probability of attending a highly competitive college likely because they would not

[28]This is a weighted average of enrollment across all of the in-state public four-year colleges, where the value of the scholarship ranged from $1,417-$1,714 at U. Mass. campuses and $910-$1,030 at state colleges.

Table 3.7: *Heterogeneity by Student Characteristics*

	Enrolled immediately, Adams college (1)	Enrolled immediately, four-year college (2)	Enrolled immediately, highly competitive (3)	On campus in year 4, four-year college (4)	Graduated within 4, four-year college (5)
(A) By district selectivity					
Eligible * bottom quintile	0.077***	0.050***	0.014*	0.010	0.010
	(0.016)	(0.011)	(0.007)	(0.012)	(0.012)
Eligible * middle quintiles	0.068***	0.011	-0.026***	-0.026***	-0.018**
	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)
Eligible * top quintile	0.015	0.004	-0.009	0.004	0.017
	(0.009)	(0.008)	(0.014)	(0.011)	(0.012)
p (Bottom = Middle)	0.609	0.005	0.000	0.016	0.066
p (Top = Middle)	0.000	0.558	0.276	0.027	0.022
(B) By race/ethnicity					
Eligible * white	0.054***	0.008	-0.026***	-0.020***	-0.009
	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)
Eligible * non-white	0.133***	0.063***	0.005	0.019	-0.024
	(0.019)	(0.019)	(0.016)	(0.021)	(0.019)
p (White = Non-white)	0.000	0.007	0.079	0.069	0.480
(C) By poverty status					
Eligible * nonpoor	0.055***	0.009	-0.027***	-0.016**	-0.010
	(0.007)	(0.006)	(0.008)	(0.006)	(0.007)
Eligible * poor	0.105***	0.037**	0.003	-0.030	-0.015
	(0.017)	(0.017)	(0.011)	(0.018)	(0.015)
p (Poor = Non-poor)	0.006	0.126	0.024	0.452	0.769
(D) By gender					
Eligible * male	0.053***	0.011	-0.030***	-0.009	-0.008
	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)
Eligible * female	0.067***	0.011	-0.017*	-0.026***	-0.014
	(0.009)	(0.008)	(0.009)	(0.009)	(0.010)
p (Male = Female)	0.258	0.958	0.303	0.159	0.679
N	88,152	88,152	88,152	88,152	88,152

Notes: Heteroskedasticity robust standard errors clustered by distance from the threshold are in parentheses (* p<.10 ** p<.05 *** p<.01). The sample consists of the high school classes of 2005-08. Each panel's baseline specification uses local linear regression with a triangular kernel of bandwidth 12. Each panel then fully interacts that baseline specification with indicators for the given categories. Panel (A) divides school districts into quintiles by the fraction of 2004 graduates who enrolled in colleges defined by Barron's as highly competitive. Panel (B) divides students into white and non-white, the latter defined by being black or Hispanic. Panel (C) divides students into non-poor and poor, the latter defined by receipt of free or reduced price lunch status. Panel (D) divides students by gender. Below each panel is the p-value from a test of the equality of the two listed coefficients.

have applied to or gained admission to such institutions. In the absence of the scholarship, these students would be attending four-year colleges of similar quality to the Adams colleges, or none at all. For these students, eligibility has no statistically significant impact on persistence and graduation on persistence, with point estimates that are slightly positive. The Adams Scholarship substantially increases college enrollment for such students, has little impact on college quality, and little impact on persistence and graduation.

Students from the top quintile districts barely react to this aid at all. There is no statistically significant evidence that they enroll in Adams colleges at higher rates, forgo highly competitive colleges, or persist or graduate at lower rates as a result of scholarship eligibility. Such students do not react presumably because they are wealthier on average and because their alternative college options are so much higher quality than the Adams colleges that the scholarship is insufficient incentive to switch.

Students from the middle quintile districts do react strongly to the aid. Eligibility raises enrollment in Adams colleges by nearly seven percentage points. Little or none of this comes from students enrolling in four-year colleges who would not have otherwise. Strikingly, two-fifths (.026/.068) of those marginal students who switch into Adams colleges do so by forgoing highly competitive colleges, unlike students from bottom or top districts. And, unlike students from bottom or top districts, only these students have clearly lower persistence and graduation rates as a result of scholarship eligibility. Unlike students from the lowest scoring districts, who are induced by the Adams Scholarship to switch sectors but not college quality level and have no graduation effects, students from middle districts are induced to forgo college quality and are the only students whose graduation rates suffer. This strengthens the case that college quality explains the scholarship's negative graduation effect.

The remaining three panels in Table 3.7 explore heterogeneity by student race/ethnicity, poverty and gender. Scholarship eligibility has a substantially stronger impact on non-white students' enrollment decisions than on white students' enrollment decisions. For non-white students, eligibility increases enrollment in an Adams college by over 13 percentage points, half of whom are students who would not otherwise have enrolled in a four-year college at all. For white students, eligibility increases Adams college enrollment by only five percentage points, nearly all of whom would have otherwise have enrolled in another four-year college. White students are the only ones who forgo more competitive colleges and see their persistence rates suffer. Non-white students do not forgo more competitive colleges and do not see a drop in their persistence rates. Roughly similar patterns are seen with poor and non-poor students. There are no statistically significant differences by gender. Taken as a whole, these results suggest that the observed negative impacts on graduation rates are concentrated in subgroups of students who are induced to forgo college quality by scholarship eligibility.

Having shown that scholarship eligibility both induced students to forgo college quality and lowered their graduation rates, in Table 3.8 we directly estimate the impact of college quality on those graduation rates. For such estimates to be valid, the exclusion restriction must hold, namely that scholarship eligibility affects graduation rates only through the college quality channel. We consider two potential violations of this exclusion restriction. First, scholarship eligibility may affect not only marginal students but infra-marginal ones as well. Our estimates suggest that roughly three-fourths of scholarship users would have attended Adams colleges in the absence of the scholarship.[29] If the financial aid were changing their graduation rates, the IV estimates would confound that channel with the quality channel. We believe this is unlikely both because the amount of money involved here is small relative to the costs of college and because that small amount of additional aid should, if anything, help students graduate by allowing them not to work while on campus. If the graduation rates of infra-marginal students were improved by this aid, the coefficients below would actually underestimate the impact of college quality on the graduation rate of marginal students.

A second potential violation of the exclusion restriction could occur if the scholarship changed factors other than college quality for the marginal students. If, for example, switching to an Adams college and remaining in state increased the probability of living at home, our estimates might confound that channel with the college quality channel. We find that story unlikely as well, given that our effects are being driven largely by students attending the U. Mass. Amherst campus in western Massachusetts, which for most students is at least an hour's drive from home. We cannot, however, definitely rule out such violations of the exclusion restriction and, as such, present the calculations below only as suggestive estimates of the impact of college quality on

[29]The scholarship raised enrollment in in-state public colleges by seven percentage points from a base of 24 percentage points, as seen in Table 3.4. Calculations using tuitions instead of enrollment yield a similar ratio.

graduation rates.

To use scholarship eligibility as an instrument for the different measures of college quality listed in each column, we run the following IV and first-stage equations:

$$GraduateIn4_{ijt} = \beta_0 + \beta_1 College\hat{Q}uality_{ijt} + \beta_2 Gap_{ijt} + \beta_3 Gap_{ijt} \times Adams_{ijt} + \epsilon_{ijt} \tag{3.4}$$

$$CollegeQuality_{ijt} = \alpha_0 + \alpha_1 Adams_{ijt} + \alpha_2 Gap_{ijt} + \alpha_3 Gap_{ijt} \times Adams_{ijt} + \nu_{ijt} \tag{3.5}$$

The first row of Table 3.8 provides the first stage coefficients by replicating the estimates seen in previous tables of the impact of scholarship eligibility on the given measure of college quality. The second row provides reduced form estimates by replicating the impact of scholarship eligibility on graduation rates from Tables 3.4. The third row contains the instrumental variables estimates themselves, the ratios of the reduced form estimates to the first stage estimates. The final row shows the OLS estimate of the same relationship without using the instrument. For each sample, the first column measures the impact of attending an Adams college for all students, while the second column conditions the sample on those immediately enrolling in a four-year college. The third and fourth columns use institutional graduation rates and our quality index as quality measures, also conditioning the sample on those immediately enrolling in a four-year college for whom such measures are observed.

The magnitudes of the IV estimates are large. For the marginal student induced by the scholarship to attend in-state public college, attending such a college lowered the probability of graduating in six years by a remarkable 36 percentage points in the 2005-06 sample, or 27 percentage points when the sample is limited to enrollers. In the 2005-08 sample, attending an Adams college lowered four-year graduation rates by 17 percentage points, or 27 percentage points for enrollers. The IV estimates in columns 2 and 6 are similar in magnitude to and statistically indistinguishable from their OLS counterparts. The coefficients in columns 3 and 7 suggest that, for these marginal students, attending a college with a four-year graduation rate one percentage point higher would translate into a roughly 1.6 percentage point increase in graduation probabilities. Differences in college-level graduation rates translate more than one-for-one into individual-level graduation rates for this subset of students, although a value of one is well within the confidence intervals of these estimates and both IV estimates are statistically indistinguishable from their

Table 3.8: *Impact of College Quality on Graduation Rates*

	Classes of 2005-06 Y = Graduation within 6 years				Classes of 2005-08 Y = Graduation within 4 years			
	Adams college (1)	Adams college (2)	Four-year grad. rate (3)	College quality (4)	Adams college (5)	Adams college (6)	Four-year grad. rate (7)	College quality (8)
First stage	0.069***	0.090***	-0.015***	-0.057***	0.060***	0.077***	-0.012***	-0.037**
	(0.010)	(0.013)	(0.005)	(0.022)	(0.007)	(0.008)	(0.003)	(0.017)
Reduced form	-0.025***	-0.025**	-0.025**	-0.025**	-0.010	-0.020***	-0.020***	-0.020***
	(0.009)	(0.010)	(0.010)	(0.010)	(0.007)	(0.008)	(0.008)	(0.008)
IV	-0.358***	-0.274***	1.651**	0.430**	-0.172	-0.265***	1.641***	0.557**
	(0.133)	(0.103)	(0.686)	(0.189)	(0.108)	(0.094)	(0.614)	(0.267)
OLS	0.026**	-0.177***	0.594***	0.120***	-0.033***	-0.230***	0.925***	0.178***
	(0.012)	(0.008)	(0.016)	(0.005)	(0.011)	(0.009)	(0.012)	(0.005)
N	41,190	29,639	29,639	29,639	88,152	63,584	63,584	63,584

Notes: Heteroskedasticity robust standard errors clustered by 12th grade school district are in parentheses (* p<.10 ** p<.05 *** p<.01). Columns 1-4 contain the high school classes of 2005-06 and columns 5-8 contain the classes of 2005-08. Each column presents an instrumental variables regression of graduation within four or six years on the endogenous regressor listed at the top of each column, where the endogenous regressor has been instrumented with aid eligibility. Each coefficient on aid eligibility is generated by local linear regression with a triangular kernel of bandwidth 12 points. The first three rows show the first stage, reduced form and instrumental variables estimates respectively. The final row shows the OLS estimate of the same regressions without using the instrument.

OLS counterparts. Finally, columns 4 and 8 suggest that attending a college of one standard deviation higher quality raises the probability of graduating by 43-56 percentage points. This is roughly three times larger than the effect estimated in Long (2008) by OLS and by instrumenting college quality by the average quality of nearby colleges.[30] That these last IV estimates are so large suggests either that omitted variable bias is causing OLS estimates to understate quality effects, that our measures of college quality understate the true differences in quality between Adams and non-Adams colleges, or that the marginal student induced to switch college due to scholarship eligibility is more sensitive to college quality than the average student.

3.6 Discussion

Our findings are consistent with Peltzman's theoretical prediction that in-kind subsidies can actually reduce consumption of the subsidized good for some individuals. Our heterogeneity results indicate that this is particularly true for students from school districts in the middle of the college enrollment distribution. The increased subsidy provided by the Adams Scholarship induced a number of such students to enroll in the public sector, where they would be spending substantially less money on tuition and, relatedly, the institutions would be spending less money on instruction. In this sense, the program achieved one of its primary goals, to draw highly skilled students into the public colleges. It is unclear, however, whether students were aware that, by choosing institutions with fewer resources to spend on instruction, they were essentially induced by the scholarship to purchase less education than they otherwise would have.

Our estimates make it difficult to explain students' enrollment decisions through a classical human capital model in which the benefits and costs of various educational options are being weighed. According to our calculations based on the American Community Survey in Massachusetts, the lifetime earnings difference between those holding only B.A.s and those with only some college is slightly under $1,000,000. Reducing one's probability of graduating by about 27 percentage points, as happened to marginal students using the scholarship, would therefore result in a $270,000 expected lifetime earnings penalty. Separate from the graduation margin, Black and Smith (2006) estimate that a one standard deviation decrease in college quality is associated with a

[30] See the first row of Table 6 in that paper.

4.2% decrease in earnings, or about $100,000 for Massachusetts B.A. holders with average lifetime earnings of $2.5 million. Attending a college of 0.5 standard deviations lower quality, as do marginal students here, thus results in an $50,000 expected lifetime earnings penalty. Either of these penalties on its own, and both together, far outweigh the value of the tuition waiver, which is at most worth less than $7,000. As such, many of these students' decisions likely failed a simple benefit-cost analysis.

We take this as strong evidence that these students did not fully understand the role of college quality in their enrollment decision. To be clear, we think that students likely understood the value of the scholarship itself, or at least were not overestimating its value. The scholarship's large and sustained enrollment impact over multiple cohorts suggest that students did not simply misunderstand the letter's promise of "four years of free tuition." Students' strong enrollment reaction have have been driven in part by the excitement of receiving aid with a formal name attached, as documented in Avery and Hoxby (2004). It is, of course, possible that some students were so financially constrained or had such high discount rates that switching into scholarship eligible institutions was a rational decision. Nonetheless, we find it more plausible that the marginal student did not understand that the role of college quality in this decision, either because she was unaware of quality differences between her college options or because she was aware but did not believe that such quality differences would affect her own graduation probability or labor market outcomes.

This work is unable to determine which, if either, of these two possibilities best explains the patterns observed here. Such distinction may be important, given that these two explanations have potentially different policy implications. If students properly understand the importance of college quality but are uninformed about actual quality differences, efforts to make such information more readily accessible may be fruitful. There has already been some movement on this front, with the 2013 release by the federal government of the College Scorecard website allowing students to quickly find and compare the graduation rates of various postsecondary institutions. If instead students are well-informed about such quality measures but discount their importance, efforts to make such measures more salient may improve college choices. The College Scorecard website implicitly does this by focusing students on a small number of measures, including graduation rates. We are, however, unaware of any clear empirical evidence on whether making college quality measures more readily available or more salient effects students' enrollment decisions.

This particular merit aid policy likely reduces social welfare. The program's costs are not listed in budget appropriations because the tuition waivers represent not expenditures but foregone revenue. The Board of Higher Education has, however, estimated that the total annual value of the waivers is roughly $13 million.[31] Roughly three-fourths of these funds flow to infra-marginal students who would have attended in-state public colleges in the absence of the scholarship. As a result of this and the low graduation rates of in-state public colleges, the scholarship has little or no impact on the number of college graduates Massachusetts produces each year. The scholarship also reduces by about 200 students per year the number of colleges degrees earned by Massachusetts high school graduates.[32] All in all, these considerations suggest the state is spending large amounts of money for little net benefit or even net harm to its students.

The only clear positive evidence we have presented, found in our heterogeneity analysis, is that this scholarship substantially increased college enrollment rates for the most disadvantaged students. Low income students, non-white students and students from the least college-oriented school districts saw their enrollment rates in four-year colleges increase by four to six percentage points. Such increased enrollment did not, however, appear to translate into increased graduation rates. More importantly, such students comprise an extremely small fraction of the total pool of those eligible for this merit aid, because of the strong negative relationship between those characteristics and standardized test scores. Most of the merit aid flowed to students who would have enrolled in college even without the scholarship.

3.7 Conclusion

We find that a relatively small amount of financial aid induces a large number of high-skilled students in Massachusetts to enroll in in-state public colleges. Many of these students forgo the opportunity to attend higher quality colleges and, in doing so, lower their own graduation rates.

[31] This estimate was communicated to us in a phone call. Our own calculations based only on the observed enrollment of Adams eligible students suggests the annual costs are closer to $25 million. Assuming the state's number is correct, this large difference is likely generated by students who do not collect their scholarships due to failure to notify their colleges of the award, failure to file a FAFSA, or failure to maintain the necessary minimum GPA.

[32] These calculations assume the local average treatment effect estimated in Table 3.4 applies to the entire population of about 15,000 Adams Scholarship recipients each year.

We argue that this is some of the clearest evidence to date that college quality likely plays an important role in determining whether students complete their degrees. This also provides a clear example of the theoretical prediction in Peltzman (1973) that in-kind subsidies of public institutions can reduce consumption of the subsidized good.

Our results highlight the importance of improving postsecondary institutions whose completion rates are low. Whether college quality operates through access to coursework, campus resources, peer effects or other channels is beyond the scope of this paper. Deeper exploration of the institution-level factors preventing college completion is needed, as this work suggests that student characteristics alone are insufficient to explain the low rates of college completion currently observed in the U.S.

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, S. Dynarski, T. J. Kane, and P. Pathak (2011). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *Quarterly Journal of Economics 126*(2), 699–748.

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.

Abdulkadiroğlu, A., J. D. Angrist, S. R. Cohodes, S. M. Dynarski, J. Fullerton, T. J. Kane, and P. Pathak (2009, January). Informing the debate: Comparing Boston's charter, pilot, and traditional schools. Technical report, The Boston Foundation, Boston, MA.

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.

Andrews, R. J., J. Li, and M. F. Lovenheim (2012). Quantile treatment effects of college quality on earnings: Evidence from administrative data in Texas. Working Paper 18068, National Bureau of Economic Research.

Angrist, J. D. (2001). Estimation of limited dependent variable models with dummy endogenous regressors. *Journal of Business & Economic Statistics 19*(1).

Angrist, J. D., S. R. Cohodes, S. M. Dynarski, J. Fullerton, T. J. Kane, P. A. Pathak, and C. R. Walters (2011). Student achievement in Massachusetts' charter schools. Center for Education Policy Research at Harvard University, Cambridge, MA.

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.

Angrist, J. D., P. A. Pathak, and C. R. Walters (2013). Explaining charter school effectiveness. *American Economic Journal: Applied Economics 5*(4), 1–27.

Avery, C. and C. M. Hoxby (2004, September). *Do and Should Financial Aid Packages Affect Students' College Choices?*, pp. 239–302. University of Chicago Press.

Bailey, M. J. and S. M. Dynarski (2011, December). Gains and gaps: Changing inequality in u.s. college entry and completion. Working Paper 17633, National Bureau of Economic Research.

Barlevy, G. and D. Neal (2012). Pay for percentile. *American Economic Review 102*(5), 1805–1831.

Bassok, D., M. Fitzpatrick, and S. Loeb (2012, December). Does state preschool crowd-out private provision? the impact of universal preschool on the childcare sector in oklahoma and georgia. Working Paper 18605, National Bureau of Economic Research.

Belley, P. and L. Lochner (2007). The changing role of family income and ability in determining educational achievement. *Journal of Human Capital 1*(1), 37–89.

Bettinger, E. P., B. T. Long, P. Oreopoulos, and L. Sanbonmatsu (2012). The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *The Quarterly Journal of Economics 127*(3), 1205–1242.

Betts, J. R. (2011). The economics of tracking in education. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), *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.

Black, D. and J. Smith (2004). How robust is the evidence on the effects of college quality? Evidence from matching. *Journal of Econometrics 121*(1), 99–124.

Black, D. and J. Smith (2006). Estimating the returns to college quality with multiple proxies for quality. *Journal of Labor Economics 24*(3), 701–728.

Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal 42*(2), 231–268.

Bound, J., M. Lovenheim, and S. Turner (2010). Why have college completion rates declined? An analysis of changing student preparation and collegiate resources. *American Economic Journal. Applied Economics 2*(3), 129.

Bound, J. and S. Turner (2007). Cohort crowding: How resources affect collegiate attainment. *Journal of Public Economics 91*(5), 877–899.

Bowen, W., M. Chingos, and M. McPherson (2009). *Crossing the finish line: Completing college at America's public universities*. Princeton University Press.

Brand, J. and C. Halaby (2006). Regression and matching estimates of the effects of elite college attendance on educational and career achievement. *Social Science Research 35*(3), 749–770.

Brewer, D., E. Eide, and R. Ehrenberg (1999). Does it pay to attend an elite private college? Cross-cohort evidence on the effects of college type on earnings. *Journal of Human Resources 34*(1), 104–123.

Brown, J. R. and A. Finkelstein (2008). The interaction of public and private insurance: Medicaid and the long-term care insurance market. *The American Economic Review 98*(3), 1083–1102.

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.

Calcagno, J., T. Bailey, D. Jenkins, G. Kienzl, and T. Leinbach (2008). Community college student success:What institutional characteristics make a difference? *Economics of Education Review 27*(6), 632–645.

Caloncio, S., M. D. Cattaneo, and R. Titiunik (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.

Cellini, S. R. (2009). Crowded colleges and college crowd-out: The impact of public subsidies on the two-year college market. *American Economic Journal: Economic Policy 1*(2), 1–30.

Center for Research on Education Outcomes (2009, June). Multiple choice: Charter school performance in 16 states. Technical report, Stanford, CA.

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.

Chevalier, A. and G. Conlon (2003). Does it pay to attend a prestigious university? Working Paper 848, Institute for the Study of Labor (IZA).

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.

Cornwell, C., D. B. Mustard, and D. J. Sridhar (2009). The enrollment effects of merit-based financial aid: Evidence from Georgia's HOPE program. *Journal of Labor Economics 24*(4), 761–786.

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.

Cutler, D. M. and J. Gruber (1996). Does public insurance crowd out private insurance? *The Quarterly Journal of Economics 111*(2), 391–430.

Dale, S. and A. Krueger (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *Quarterly Journal of Economics 117*(4), 1491–1527.

Dale, S. and A. B. Krueger (2011). Estimating the return to college selectivity over the career using administrative earnings data. Working Paper 17159, National Bureau of Economic Research.

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. and S. Dynarski (2010). College aid. In *Targeting investments in children: Fighting poverty when resources are limited*, pp. 283–302. University of Chicago Press.

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 (2011). Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone. *American Economic Journal: Applied Economics 3*(3), 158–187.

Dobbie, W. and R. G. Fryer (2013). Getting beneath the veil of effective schools: Evidence from New York City. *American Economic Journal: Applied Economics 5*(4), 28–60.

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.

Dobbie, W. and R. G. Fryer (Forthcoming). The medium-term impacts of high-achieving charter schools. *Journal of Political Economy*.

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. (2000). Hope for whom? Financial aid for the middle class and its impact on college attendance. *National Tax Journal 53*(3), 629–661.

Dynarski, S. (2008). Building the stock of college-educated labor. *Journal of Human Resources 43*(3), 576–610.

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.

Fitzpatrick, M. D. and D. Jones (2012). Higher education, merit-based scholarships and post-baccalaureate migration. Working Paper 18530, National Bureau of Economic Research.

Forstan, K., N. Verbitsky-Savitz, E. Kopa, and P. Gleason (2012, April). Using experimental evaluation of charter schools to test whether nonexperimental comparison group methods can replicate experimental impact estimates. Technical Report NCCE Technical Methods Report 2012-4019, Institute of Education Sciences, Washington, D.C.

Ganderton, P. T. (1992). The effect of subsidies in kind on the choice of a college. *Journal of Public Economics 48*(3), 269–292.

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.

Gleason, P., M. Clark, C. C. Tuttle, and E. Dwoyer (2010, June). The evaluation of charter school impacts: Final report. Technical Report NCEE 2010-4029, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Goodman, J. (2008). Who merits financial aid?: Massachusetts' Adams Scholarship. *Journal of Public Economics 92*(10-11), 2121–2131.

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.

Hamilton, L. (2003). Assessment as a policy tool. *Review of Research in Education 27*, 25–68.

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., S. Murarka, and J. Kang (2009, September). How New York City's charter schools affect achievement. Technical report, New York City Charter Schools Evaluation Project, Cambridge, MA.

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.

Jacob, B. A. (2005). Accountability, incentives and behavior: the impact of high-stakes testing in the chicago public schools. *Journal of Public Economics 89*(5-6), 761–796.

Jacob, B. A. and S. D. Levitt (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics 118*(3), 843–877.

Kane, T. (2007). Evaluating the impact of the DC Tuition Assistance Grant program. *Journal of Human Resources 42*(3), 28.

Kane, T. J. (2006). Public intervention in post-secondary education. *Handbook of the Economics of Education 2*, 1369–1401.

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.

Koretz, D. M. (2008). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.

Lang, K. and E. Siniver (2011). Why is an elite undergraduate education valuable? Evidence from Israel. *Labour Economics 18*(6), 767–777.

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

Leigh, D. and A. Gill (2003). Do community colleges really divert students from earning bachelor's degrees? *Economics of Education Review 22*(1), 23–30.

Long, B. T. (2004, August). Does the format of a financial aid program matter? The effect of state in-kind tuition subsidies. *Review of Economics and Statistics 86*(3), 767–782.

Long, M. (2008). College quality and early adult outcomes. *Economics of Education Review 27*(5), 588–602.

Loury, L. and D. Garman (1995). College selectivity and earnings. *Journal of Labor Economics 13*(2), 289–308.

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

Massachusetts Department of Elementary and Secondary Education (2000, November). Massachusetts mathematics curriculum framework. Available: http://www.doe.mass.edu/frameworks/math/2000/toc.html.

Massachusetts Department of Elementary and Secondary Education (2007). 2007 MCAS technical report. Available: http://www.mcasservicecenter.com/documents/MA/Technical

Massachusetts Department of Elementary and Secondary Education (2011a). Adequate yearly progress: 2011 lists of Massachusetts schools and districts by NCLB accountability status and accountability and assistance level.

Massachusetts Department of Elementary and Secondary Education (2011b). Item by item results.

Mcclain, M. 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.

McMurrer, J. (2007). Choices, changes, and challenges: Curriculum and instruction in the NCLB era. Technical report, Center on Education Policy, Washington, D.C.

Merseth, K. K. (2010). High-performing charter schools: Serving two masters? In *Hopes, fears, and reality: A balanced look at American charter schools in 2009*. Center on Reinventing Public Education, University of Washington Bothell.

Merseth, K. K., K. Cooper, J. Roberts, M. C. Tieken, J. Valant, and C. Wynne (2009). *Inside urban charter schools: Promising practices and strategies in five high-performing schools*. Cambridge, MA: Harvard Education Press.

Muralidharan, K. and V. Sundararaman (2011, February). Teacher performance pay: Experimental evidence from India. *Journal of Political Economy 119*(1), 39–77.

National Mathematics Advisory Panel (2008). Foundations for success: Report of the National Mathematics Advisory Panel. Technical report, U.S. Department of Education, Washington, D.C.

Neal, D. and D. W. Schanzenbach (2010). Left behind by design: Proficiency counts and test-based accountability. *Review of Economics and Statistics 92*(2), 263–283.

Nichols, S. L. and D. C. Berliner (2007). *Collateral Damage: How High-Stakes Testing Corrupts America's Schools*. Harvard Education Press.

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

Pallais, A. (2009). Taking a chance on college is the Tennessee education lottery scholarship program a winner? *Journal of Human Resources 44*(1), 199–222.

Papay, J. P., R. J. Murnane, and J. B. Willett (2010, March). The consequences of high school exit examinations for low-performing urban students: Evidence from Massachusetts. *Educational Evaluation and Policy Analysis 32*(1), 5–23.

Papay, J. P., J. B. Willett, and R. J. Murnane (2011a). Extending the regression-discontinuity approach to multiple assignment variables. *Journal of Econometrics 161*, 203–207.

Papay, J. P., J. B. Willett, and R. J. Murnane (2011b). High-school exit examinations and the schooling decisions of teenagers: A multi-dimensional regression-discontinuity approach. Working Paper 17112, National Bureau of Economic Research.

Peltzman, S. (1973). The effect of government subsidies-in-kind on private expenditures: The case of higher education. *The Journal of Political Economy 81*(1), 1–27.

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.

Reardon, S. F. and J. P. Robinson (2012). Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness 5*(1), 83–104.

Reynolds, C. L. (2012). Where to attend? Estimating the effects of beginning college at a two-year institution. *Economics of Education Review 31*(4), 345–362.

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.

Rouse, C. (1995). Democratization or diversion? The effect of community colleges on educational attainment. *Journal of Business & Economic Statistics 13*(2), 217–224.

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

Sandy, J., A. Gonzalez, and M. Hilmer (2006). Alternative paths to college completion: Effect of attending a 2-year school on the probability of completing a 4-year degree. *Economics of Education Review 25*(5), 463–471.

Stange, K. (2009). Ability sorting and the importance of college quality to student achievement: Evidence from community colleges. *Education Finance and Policy 7*(1), 1–32.

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.

The Boston Globe (2011). 2011 MCAS results.

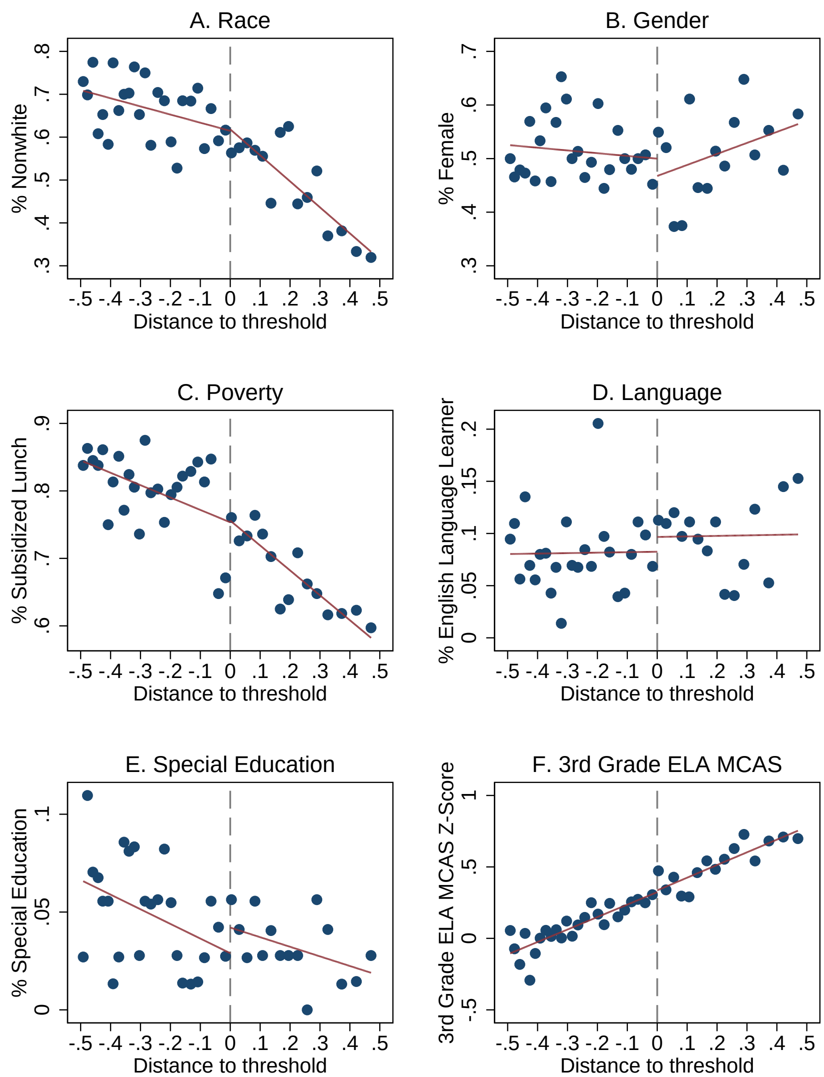
Zimmer, R., B. Gill, K. Booker, S. Lavertu, T. R. Sass, and J. Witte (2009). *Charter schools in eight states: Effects on achievements, attainment, integration, and competition*. Santa Monica, CA: Rand Corporation.

Appendix A

Appendix to Chapter 1

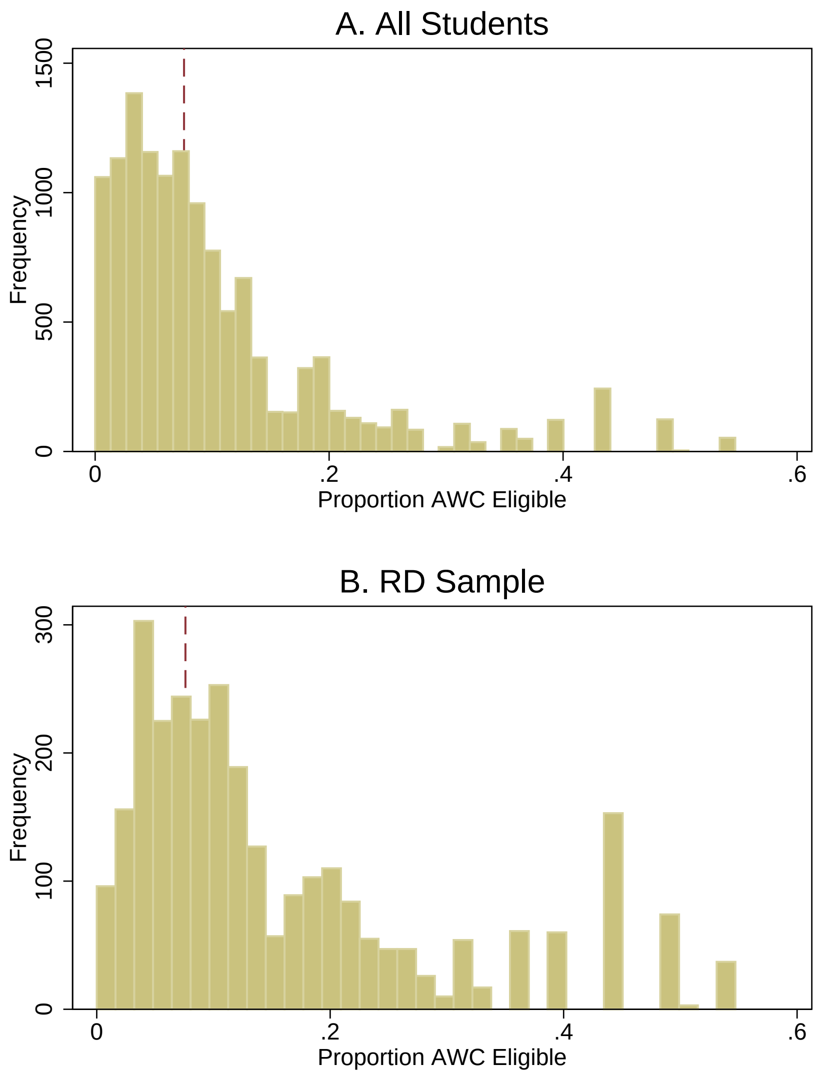
A.1 Supplemental Tables and Figures

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 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 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 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

Notes: 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

Notes: 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 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 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 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 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 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 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 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 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 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 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 1.6 (for high school outcomes), and Table 1.6 (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.

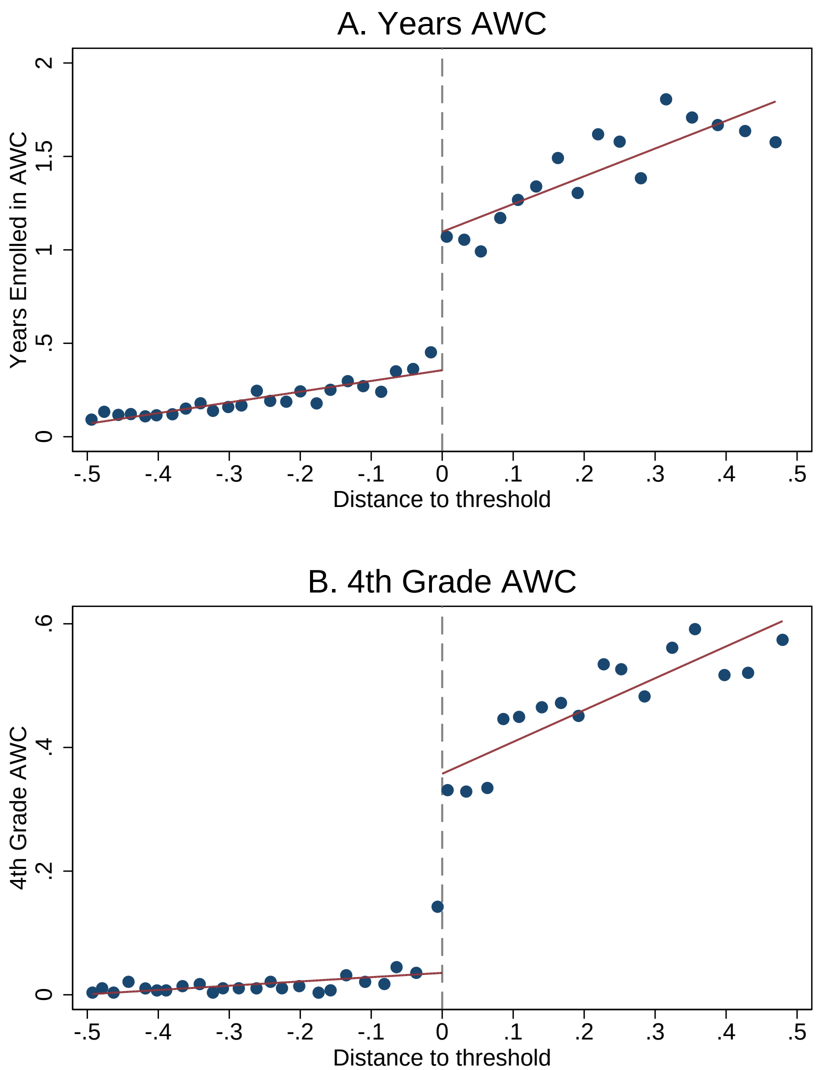
Table A.14: *Descriptive Statistics for the Regression Discontinuity Sample, by 3rd Grade School Type*

	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 eligibility rate greater than or equal to 7.6 percent. Column (6) restricts this sample further to those whose 3rd grade school has an AWC eligibility rate below to 7.6 percent.

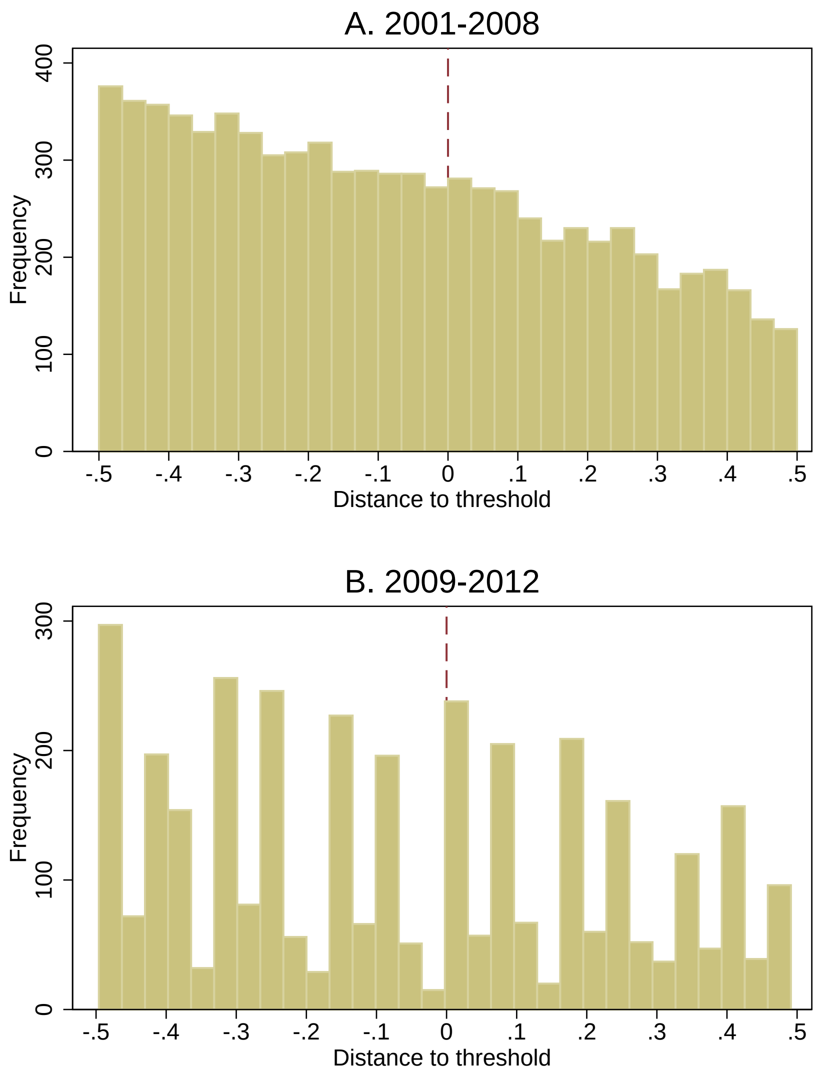
A.2 Main Results Using All Available 3rd Grade Cohorts

Figure A.3: *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 A.4: *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 A.15: *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

Notes: 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 A.16: *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

Notes: 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 A.17: *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 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 A.18: *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 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 A.19: *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 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 A.20: *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 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 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.21: *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 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 A.22: *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 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.

Appendix B

Appendix to Chapter 2

B.1 Supplemental Tables and Figures

Table B.1: *Outcome Years*

	Math			ELA			Science
	Grade 6	Grade 7	Grade 8	Grade 6	Grade 7	Grade 8	Grade 8
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rare Items Sample	2007-2011	2007-2011	2007-2011	2007-2011	2007-2011	2007-2011	2007-2011
Full Sample	2004-2011	2006-2011	2006-2011	2006-2011	2006-2011	2006-2011	2006-2011

Notes: Years indicate the spring of the school year, when the MCAS is administered. Information on the standards associated with each item was first published in 2007, thus the limited years for the rare items sample. The 7th grade math, 6th grade ELA, and 8th grade ELA MCAS exams were administered for the first time in spring 2006. The 6th and 8th grade math, 7th grade ELA, and 8th grade science MCAS exams were administered in years prior to those listed, however the first students that participated in the lotteries in the sample take the exam in the years noted.

Table B.2: *Covariate Balance between Charter Applicants Offered a Seat and Not Offered a Seat in Charter School Lotteries*

	Rare Standards Sample Difference (Offered - Not Offered)		Full Sample Difference (Offered - Not Offered)	
	Coefficient	S.E.	Coefficient	S.E.
	(1)	(2)	(3)	(4)
Latino/a	0.048***	(0.017)	0.044***	(0.015)
African-American	-0.038**	(0.019)	0.037**	(0.017)
White	-0.006	(0.015)	-0.006	(0.013)
Asian	-0.001	(0.006)	-0.001	(0.005)
Female	-0.021	(0.020)	-0.005	(0.019)
Free or Reduced Price Lunch	0.028	(0.018)	0.017	(0.017)
Special Education	0.004	(0.015)	0.000	(0.014)
English Language Learner	0.016	(0.012)	0.014	(0.010)
Baseline Standardized Math Score	-0.018	(0.040)	-0.018	(0.037)
Baseline Standardized ELA Score	-0.045	(0.038)	-0.033	(0.035)
Sample Size	3392		4036	
P-value from F-test	0.206		0.373	

Notes: This table reports coefficients on regressions of the variable indicated in each row on an indicator variable equal to one if the student was offered a seat at a charter through the lottery. The sample is restricted to charter school applicants without sibling priority in the lottery, who attended a public or charter charter school in their year of application, and who have baseline demographic characteristics and test scores. All regressions include , lottery risk sets, which are a set of dummies for the combination of schools applied to by year, and year of baseline and year of birth dummies. Regressions use robust standard errors. F tests are for the null hypothesis that the coefficients on winning the lottery in all regressions are all equal to zero. These tests statistics are calculated for the subsample that has non-missing values for all variables tested. Students must have at least one MCAS outcome to be included in the table. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table B.3: *First Stage, Reduced Form, and 2SLS Estimates for 6th, 7th, and 8th Grades Combined*

Subscale Outcome	First Stage (1)	Math Reduced Form (2)	2SLS (3)	First Stage (4)	ELA Reduced Form (5)	2SLS (6)
(A) Rare Standards Sample						
All Items	0.986***	0.327***	0.332***	0.988***	0.181***	0.183***
	(0.064)	(0.032)	(0.030)	(0.064)	(0.029)	(0.028)
Rare	0.986***	0.357***	0.363***	1.043***	0.129***	0.124***
	(0.064)	(0.037)	(0.035)	(0.068)	(0.034)	(0.032)
Somewhat Common	0.986***	0.331***	0.336***	0.988***	0.162***	0.164***
	(0.064)	(0.033)	(0.031)	(0.064)	(0.030)	(0.030)
Common	0.986***	0.274***	0.277***	0.988***	0.173***	0.175***
	(0.064)	(0.031)	(0.028)	(0.064)	(0.030)	(0.028)
N		6633			6600	
(B) Full Sample						
All Items	0.975***	0.358***	0.367***	0.994***	0.176***	0.177***
	(0.060)	(0.031)	(0.030)	(0.060)	(0.028)	(0.026)
Multiple Choice	0.975***	0.373***	0.383***	0.994***	0.163***	0.164***
	(0.060)	(0.032)	(0.031)	(0.060)	(0.027)	(0.026)
Short Answer	0.975***	0.349***	0.359***	-	-	-
	(0.060)	(0.035)	(0.033)	-	-	-
Open Response	0.975***	0.277***	0.284***	0.994***	0.148***	0.149***
	(0.060)	(0.030)	(0.029)	(0.060)	(0.035)	(0.034)
N		7581			7364	

Notes: This table reports first stage, reduced form, and 2SLS coefficients for regressions with 6th, 7th, and 8th grade outcomes pooled across grades. In addition to all of the model notes in Table 2.4, it includes dummies for grade level and double clusters standard errors by student and school by year. Science is excluded from this table as it is only offered in 8th grade. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table B.4: *Outcome Means in Raw Score Points*

	Not Offered a Seat in the Charter Lottery							Offered a Seat in the Charter Lottery						
	Math			ELA			Science	Math			ELA			Science
	6th	7th	8th	6th	7th	8th	8th	6th	7th	8th	6th	7th	8th	8th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(A) Rare Standards Sample														
All Items	34.74	34.09	33.20	34.26	34.43	35.20	27.77	38.25	37.58	37.15	35.24	36.29	36.29	31.05
Rare	5.68	5.32	6.70	1.21	1.40	2.38	3.44	6.07	5.96	7.48	1.13	1.40	2.36	3.68
Somewhat Common	12.41	8.40	6.96	6.63	4.83	5.42	6.63	14.21	9.12	7.60	7.20	5.19	5.49	7.13
Common	16.64	20.37	19.54	26.43	28.20	27.41	17.70	17.97	22.50	22.07	26.91	29.70	28.44	20.24
(B) Full Sample														
All Items	34.33	33.90	32.79	34.20	34.47	35.12	32.79	37.94	37.47	37.09	34.99	36.21	36.30	37.09
Multiple Choice	20.85	19.81	20.39	25.58	26.11	26.18	20.94	22.68	21.85	22.71	26.14	27.07	26.90	23.27
Short Answer	3.20	3.66	3.25	1.00	1.00	1.00	1.00	3.60	4.10	3.74	1.00	1.00	1.00	1.00
Open Response	10.28	10.42	9.15	8.62	8.36	8.94	6.88	11.66	11.52	10.63	8.84	9.14	9.40	7.97
Geometry	4.24	4.28	3.79	-	-	-	-	4.78	4.77	4.44	-	-	-	-
Measurement	4.18	3.81	3.88	-	-	-	-	4.65	4.16	4.48	-	-	-	-
Num. Sense & Operations	11.66	8.33	8.43	-	-	-	-	12.93	9.42	9.44	-	-	-	-
Patterns, Alg. & Relations	9.45	9.94	9.50	-	-	-	-	10.25	10.74	10.65	-	-	-	-
Data Analysis, Stat. & Prob.	4.80	7.53	7.20	-	-	-	-	5.34	8.38	8.08	-	-	-	-
Reading	-	-	-	29.95	30.97	31.07	-	-	-	-	30.67	32.69	32.11	-
Language and Literature	-	-	-	4.25	3.49	4.05	-	-	-	-	4.32	3.52	4.19	-
Earth and Space Science							6.96							7.74
Life Science	-	-	-	-	-	-	7.79	-	-	-	-	-	-	8.83
Physical Science	-	-	-	-	-	-	6.47	-	-	-	-	-	-	7.59
Tech. and Engineering	-	-	-	-	-	-	6.59	-	-	-	-	-	-	7.08

Notes: This table reports the mean outcome for students who did not win a seat in the charter school lottery (Columns 1-7) and students that did win a seat in the lottery (Columns 8-14), in raw score MCAS points. The difference in means roughly corresponds to the reduced form estimates in Tables 2.3 and 2.5.

Table B.5: *2SLS Estimates on Standards Categorized by Last Year's Test, Effect of Attending a Charter School, Per Year of Attendance, on MCAS Outcomes*

	Math		Science
	6th	8th	8th
Subscale Outcome	(1)	(2)	(3)
(A) Last Year's Standards Sample (2008-2011)			
All Items	0.489***	0.219***	0.269***
	(0.051)	(0.032)	(0.039)
Standards not on Last Year's Test	0.545***	0.126***	0.220***
	(0.068)	(0.039)	(0.044)
Standards on Last Year's Test	0.462***	0.224***	0.276***
	(0.049)	(0.033)	(0.039)
N	2276	1596	1595

Notes: The notes for this table are the same as those for Table 2.4, with different outcomes, defined by whether or not a standard appear on last year's test. 7th grade math and all grades of ELA tested for each standard in almost every test administration, so it is impossible to create these outcomes for those grades and subjects. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table B.6: *Match from Lottery Records to SIMS*

		Fraction with SIMS match			
Lottery cohort	Number of records (1)	Total (2)	Offered (3)	Not Offered (4)	Offered > Not Offered? (5)
2002	295	0.908	0.934	0.859	Yes
2003	302	0.861	0.873	0.804	No
2004	300	0.887	0.930	0.848	Yes
2005	678	0.934	0.968	0.883	Yes
2006	837	0.952	0.968	0.919	Yes
2007	1026	0.958	0.983	0.914	Yes
2008	1225	0.930	0.959	0.881	Yes
2009	1414	0.897	0.896	0.898	No
2010	1254	0.923	0.956	0.904	Yes
All	7331	0.924	0.947	0.894	Yes

Notes: This table summarizes the match from the lottery records to the SIMS data. The sample excludes disqualified applicants, late applicants, out-of-area applicants, and siblings. Offered > not offered determined from a two group mean comparison t-test with a p-value of .95.

Table B.7: *2SLS for Cohorts with Same Match Rates*

Subscale Outcome	Math Grade 6 (1)	ELA Grade 6 (2)
(A) Rare Standards Sample		
All Items	0.524***	0.198***
	(0.070)	(0.063)
Rare	0.516***	0.332**
	(0.061)	(0.131)
Somewhat Common	0.546***	0.181***
	(0.093)	(0.070)
Common	0.481***	0.158**
	(0.070)	(0.071)
N	695	694
(B) Full Sample		
Full Sample All Items	0.534***	0.210***
	(0.070)	(0.062)
Multiple Choice	0.558***	0.221***
	(0.073)	(0.050)
Short Answer	0.594***	-
	(0.081)	-
Open Response	0.369***	0.092
	(0.070)	(0.117)
N	767	697

Notes: The notes for this table are the same as those for Table 2.4, except here results are only for lottery applicants in 2002 and 2009, when the SIMS match rate across the offered and not offered group was not significantly different. 7th and 8th grade results are not reported due to small sample size. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table B.8: *2SLS with Imputed Outcomes for Attriters*

		Math			ELA		Science
Subscale Outcome	Grade 6 (1)	Grade 7 (2)	Grade 8 (3)	Grade 6 (4)	Grade 7 (5)	Grade 8 (6)	Grade 8 (7)
(A) Rare Standards Sample							
All Items	0.465***	0.288***	0.243***	0.184***	0.226***	0.121***	0.282***
	(0.046)	(0.038)	(0.034)	(0.042)	(0.038)	(0.033)	(0.037)
Rare	0.407***	0.349***	0.195***	0.164***	0.111**	0.064	0.193***
	(0.056)	(0.050)	(0.037)	(0.062)	(0.047)	(0.040)	(0.039)
Somewhat Common	0.528***	0.273***	0.206***	0.194***	0.177***	0.106***	0.177***
	(0.069)	(0.036)	(0.034)	(0.048)	(0.042)	(0.037)	(0.041)
Common	0.431***	0.246***	0.248***	0.153***	0.225***	0.126***	0.303***
	(0.045)	(0.037)	(0.036)	(0.044)	(0.040)	(0.033)	(0.038)
N	2963	2366	1951	2928	2372	1949	1951
(B) Full Sample							
All Items	0.526***	0.306***	0.268***	0.167***	0.218***	0.131***	0.287***
	(0.049)	(0.036)	(0.034)	(0.041)	(0.034)	(0.030)	(0.036)
Multiple Choice	0.541***	0.335***	0.264***	0.175***	0.191***	0.111***	0.285***
	(0.052)	(0.038)	(0.033)	(0.038)	(0.033)	(0.028)	(0.038)
Short Answer	0.535***	0.295***	0.248***	-	-	-	-
	(0.059)	(0.043)	(0.037)	-	-	-	-
Open Response	0.396***	0.223***	0.238***	0.096	0.199***	0.139***	0.243***
	(0.049)	(0.034)	(0.039)	(0.061)	(0.048)	(0.047)	(0.036)
N	3561	2536	2086	3237	2688	2087	2086

Notes: The notes for this table are the same as those for Table 2.4, except here baseline scores are used as the outcome for students missing outcome data. ***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table B.9: *Sample Selection*

Applications to charter schools with sufficient records that do not offer enrollment to all applicants	8183
Excluding disqualified applications (wrong grade, repeat application, etc.)	8159
Excluding late applications	8092
Excluding out-of-area applications	8018
Excluding applications with sibling priority	7331
Excluding applications not matched to state database	6771
Transforming to one observation to per applicant	5213
Excluding students without a baseline demographics	4339
Excluding students without a baseline test score in any subject	4065
Excluding students without an outcome test score in any subject or grade	3395

Table B.10: *Charter School Participation in Lottery Based Analysis*

	Available Spring Lottery Data (1)	Grade Range (2)	Notes (3)
Academy of the Pacific Rim Charter Public School	2005-2010	5-12	
Boston Collegiate Charter School	2002-2010	5-12	
Boston Preparatory Charter Public School	2005-2010	6-11	Initial offer only in 2005.
Dorchester Collegiate Academy Charter School	x	4-5	Opened September 2009.
Edward Brooke Charter School	2006-2009	K-8	Became K-8 in 2006. Initial offer only in 2006. Only middle grade entry lotteries used.
Excel Academy Charter School	2008-2010	5-8	
MATCH Charter Public School	2008-2010	6-12	Opened middle school 2008.

Notes: Schools that have entry grade lotteries only in kindergarten are excluded, which excludes Boston Community Charter School and Neighborhood House Charter School. Schools that closed in the relevant time period are excluded, which excludes Fredrick Douglass Charter School (closed 2005) and Uphams Corner Charter School (closed 2010). The remaining schools that do not contribute lotteries to the analysis are not oversubscribed or do not have sufficient lottery records.

Appendix C

Appendix to Chapter 3

C.1 Supplemental Tables and Figures

Figure C.1: *Award Letter to Class of 2005*

THE COMMONWEALTH OF MASSACHUSETTS
EXECUTIVE DEPARTMENT
STATE HOUSE ● BOSTON, MA 02133
(617) 725-4000

MITT ROMNEY
GOVERNOR

KERRY HEALEY
LIEUTENANT GOVERNOR

December 9, 2004

Dear [redacted] :

Congratulations! You are one of the first recipients of the John and Abigail Adams Scholarship. The Adams scholarship is good for **four years of free tuition** at any University of Massachusetts campus, or any state or community college. Your outstanding MCAS results automatically qualify you to receive this award.

We created this merit scholarship program to reward your hard work and achievement, and to encourage you to go to college at one of our top-notch public higher education institutions.

With the support of the Board of Higher Education, the Class of 2005 is now the first to be awarded this opportunity. It is the strongest expression we can make of our commitment to attracting students like you- the best and brightest in the state- to our Commonwealth's public higher education system.

I encourage you to apply to any of the campuses on the attached list. Congratulations again, and best wishes for your continued success.

Sincerely,

Mitt Romney

Notes: Above is a copy of the Adams Scholarship award letter sent to the first treated cohort, high school seniors in the class of 2005.

Figure C.2: *Award Letter to Class of 2006*

THE COMMONWEALTH OF MASSACHUSETTS
EXECUTIVE DEPARTMENT
STATE HOUSE • BOSTON 02133
(617) 725-4000

MITT ROMNEY
GOVERNOR

KERRY HEALEY
LIEUTENANT GOVERNOR

December 9, 2005

Dear [redacted]:

Congratulations! You are a recipient of the John and Abigail Adams Scholarship. The Adams scholarship offers **four years of free tuition** to full-time students attending any University of Massachusetts campus, or any state or community college, beginning with the fall 2006 semester. Your outstanding MCAS results have qualified you to receive this award.

We created this merit scholarship program to reward your hard work and achievement. With the support of the Board of Higher Education, the Class of 2006 is now the second class to be awarded this opportunity. It is the strongest expression we can make of our commitment to attracting students like you — the best and brightest in the state — to our Commonwealth's public higher education system.

I encourage you to read the enclosed material and apply to any of the campuses on the attached list. Please present a copy of this letter once you are accepted to the college of your choice as proof of your award. College fees and rooming costs are not included in this scholarship award, so it is in your interest to complete the Free Application for Federal Student Aid (FAFSA) to help with these costs.

Congratulations again, and best wishes for your continued success.

Sincerely,

Mitt Romney

Notes: Above is a copy of the Adams Scholarship award letter sent to the second treated cohort, high school seniors in the class of 2006.

Figure C.3: *Award Letter to Class of 2012*

Massachusetts Executive Office of Education

One Ashburton Place, Room 1403, Boston, Massachusetts 02108 *Telephone: (617) 979-8340*

PAUL REVILLE
Secretary of Education

October 2011

Dear [redacted]:

Congratulations!

You have qualified to receive a John and Abigail Adams Scholarship, which entitles you to four years of free tuition **upon your acceptance to** a participating Massachusetts public institution of higher education, including a University of Massachusetts campus, a Massachusetts state university, or a community college.

Now in its eighth year, the Adams Scholarship rewards high academic achievement on MCAS tests, and provides families of college-bound students with financial assistance. Please note that the Adams Scholarship covers **tuition only**, and does not include college fees.

Please review the enclosed guidelines carefully to determine whether you meet the eligibility requirements. If you do, I encourage you to apply to one of the campuses on the attached list.

It is **extremely important** that you **make a copy** of this letter and keep the letter and copy in a safe place. In order to receive the scholarship, **you must submit this letter** to the financial aid office of the Massachusetts public college or university to which you have been accepted and complete the online Free Application for Federal Student Aid (FASFA).

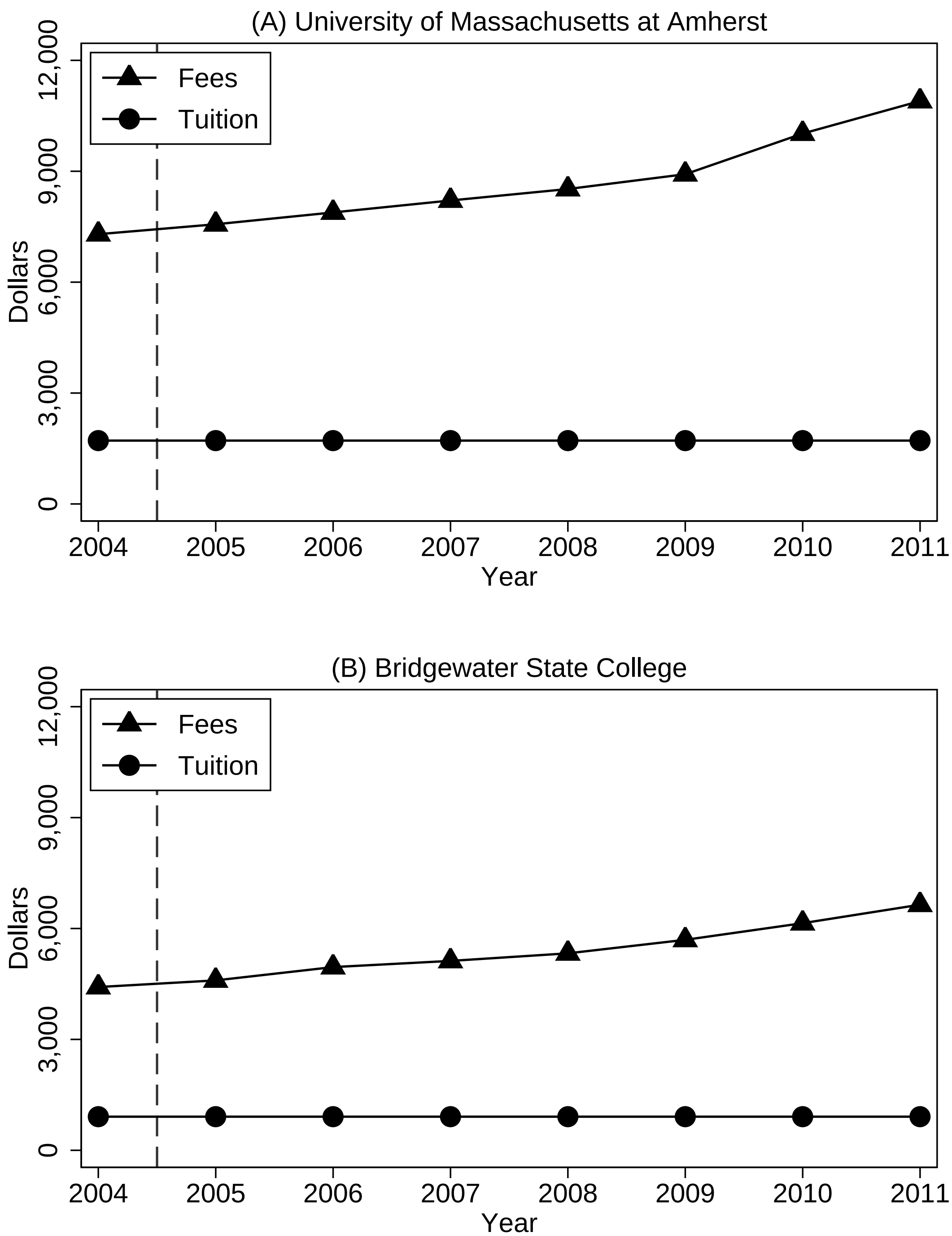
Congratulations again, and best wishes for your continued success in college and beyond.

Sincerely,

Paul Reville Secretary of Education	Mitchell D. Chester, Ed.D. Commissioner of Elementary and Secondary Education	Dr. Richard Freeland Commissioner of Higher Education

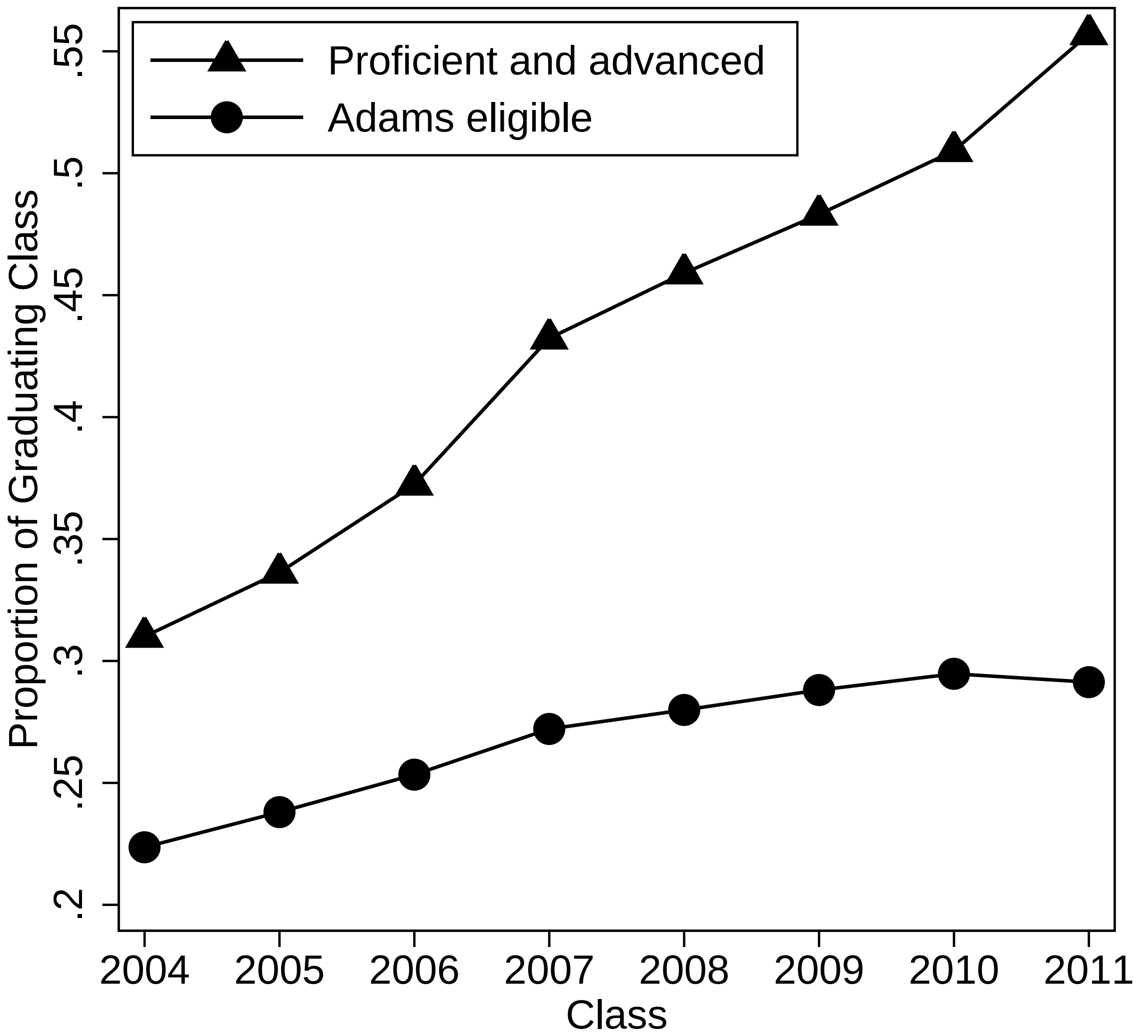
Notes: Above is a copy of the Adams Scholarship award letter sent to a recently treated cohort, high school seniors in the class of 2012.

Figure C.4: *Tuition and Fees at Two Typical Adams Colleges*



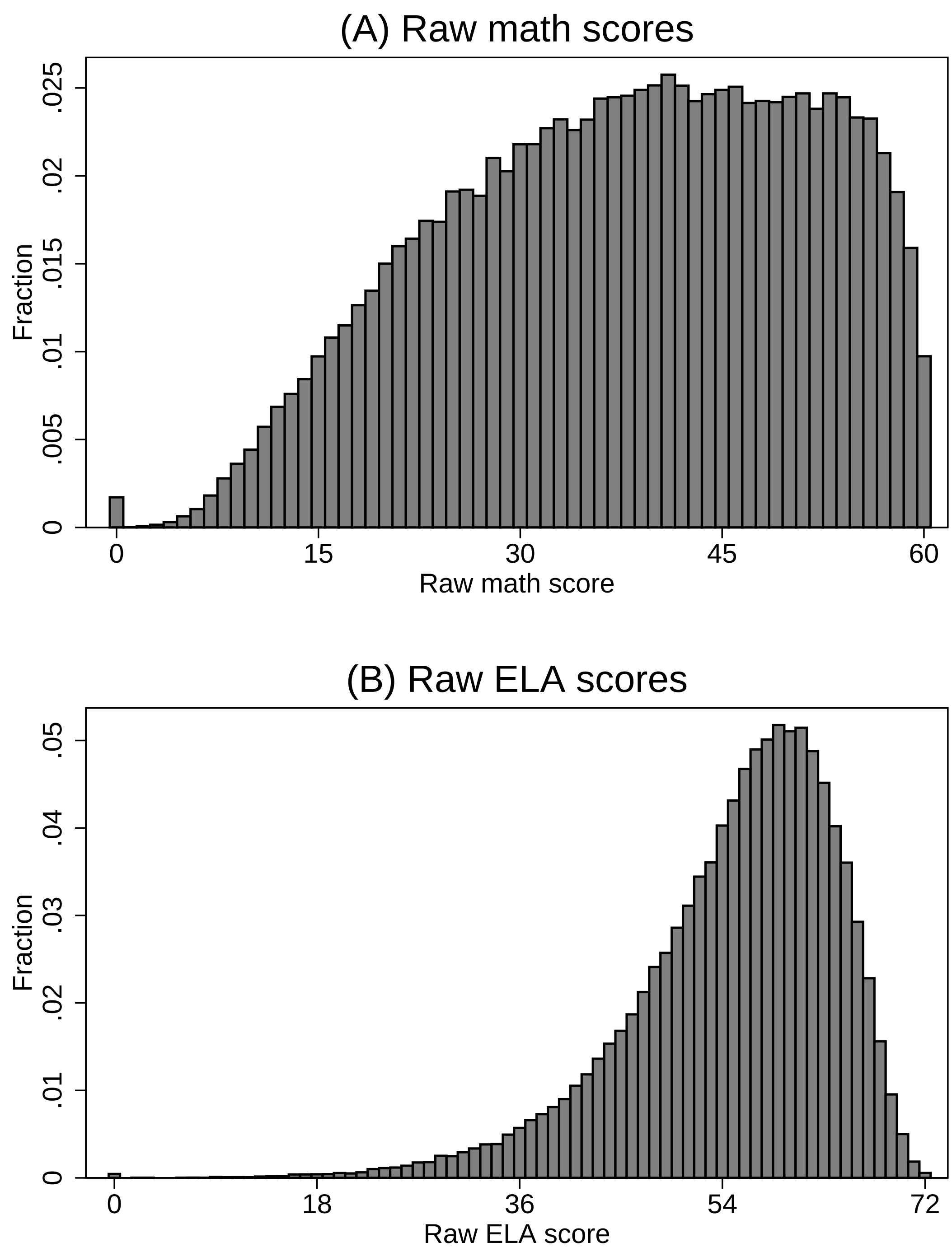
Notes: The above panels show tuition and fees over time for the largest U. Mass. campus and the largest state college. Data come from `http://www.mass.edu/campuses/tuitionfees.asp, accessed on May 28, 2013.`

Figure C.5: *Adams Eligibility by High School Class*



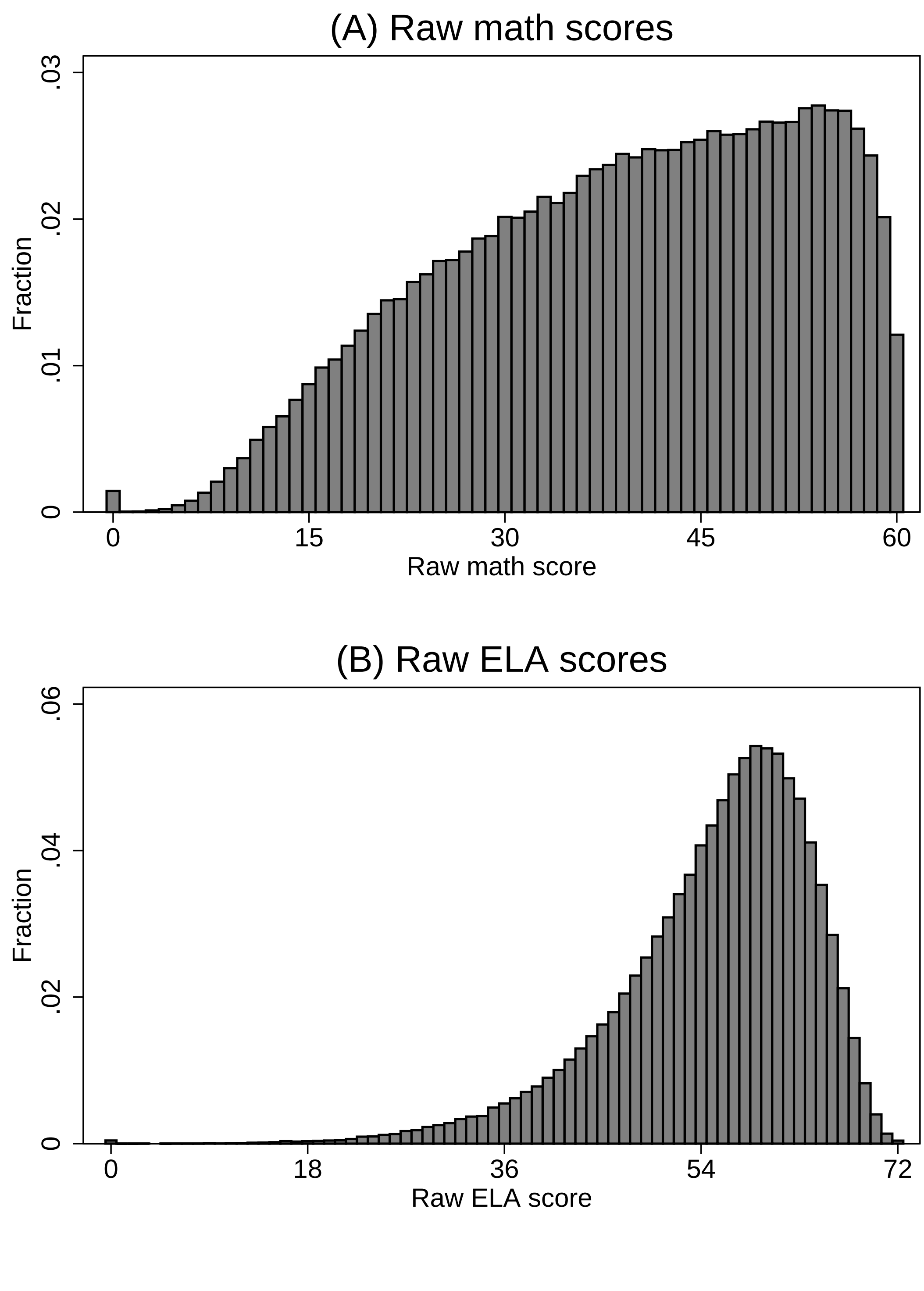
Notes: The top line shows the fraction of students scoring advanced on one MCAS section and proficient or advanced on the other. The bottom line shows the fraction of students deemed eligible for the Adams Scholarship. Calculations are based on data from DESE.

Figure C.6: *Density of Raw Scores, Classes of 2005-06*



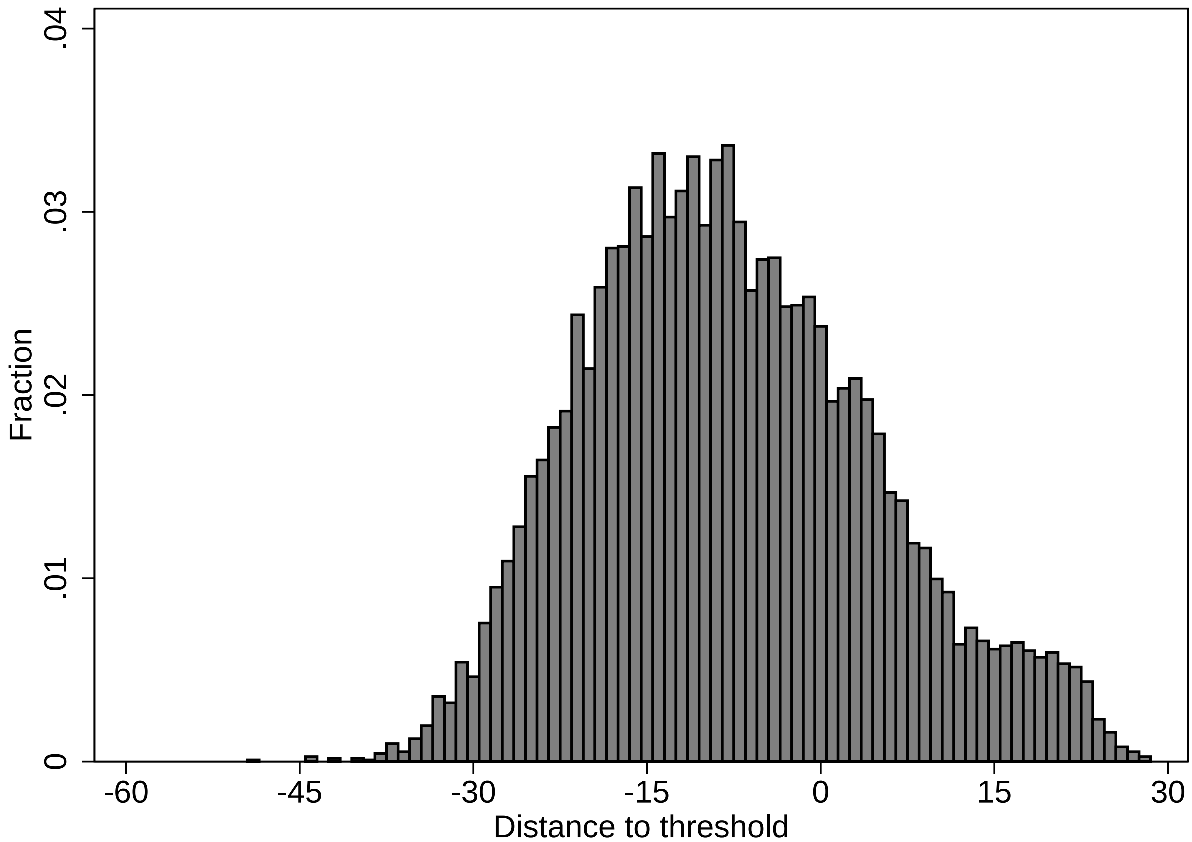
Notes: The above panels show, for the high school classes of 2005-06, the full distribution of raw math and ELA MCAS scores that underlie construction of the running variable. Calculations are based on data from DESE.

Figure C.7: *Density of Raw Scores, Classes of 2005-08*



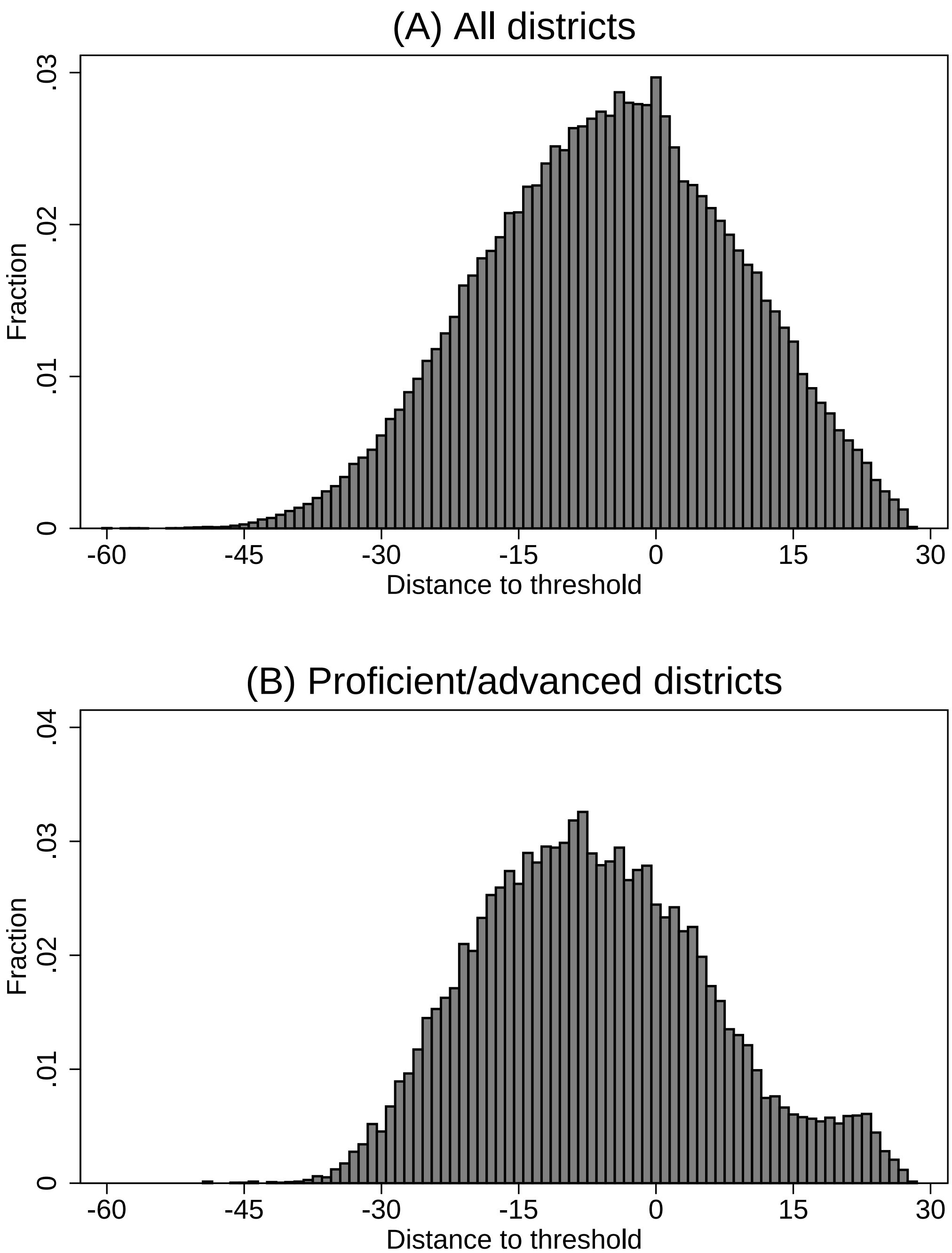
Notes: The above panels show, for the high school classes of 2005-08, the full distribution of raw math and ELA MCAS scores that underlie construction of the running variable. Calculations are based on data from DESE.

Figure C.8: *Density of the Running Variable in Proficient/Advanced Districts, Classes of 2005-06*



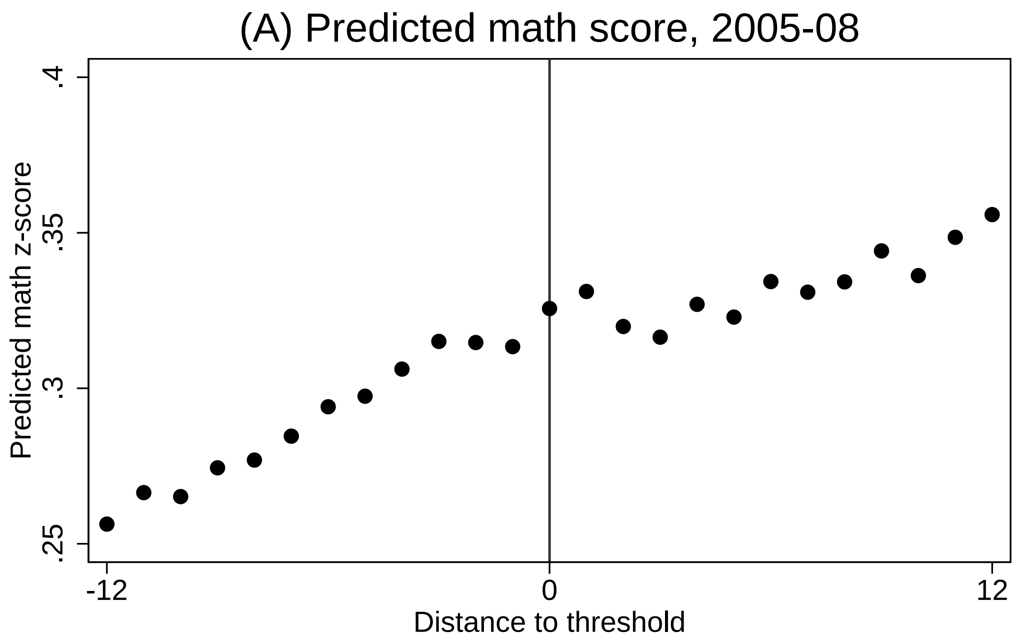
Notes: The above figure shows, for the high school classes of 2005-06, the full distribution of the running variable in school districts where the proficient/advanced threshold is binding and the top 25% threshold is irrelevant. Calculations are based on data from DESE.

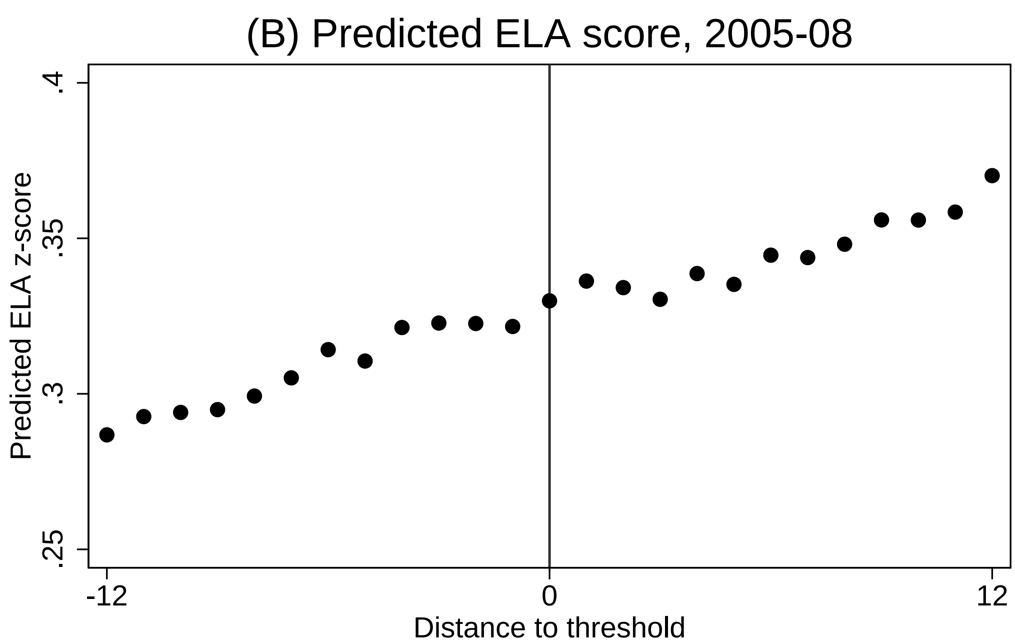
Figure C.9: *Density of the Running Variable, Classes of 2005-08*



Notes: Panel A shows, for the high school classes of 2005-08, the full distribution of the running variable in all school districts. Panel B shows the distribution of the running variable in districts where the proficient/advanced threshold is binding and the top 25% threshold is irrelevant. Calculations are based on data from DESE.

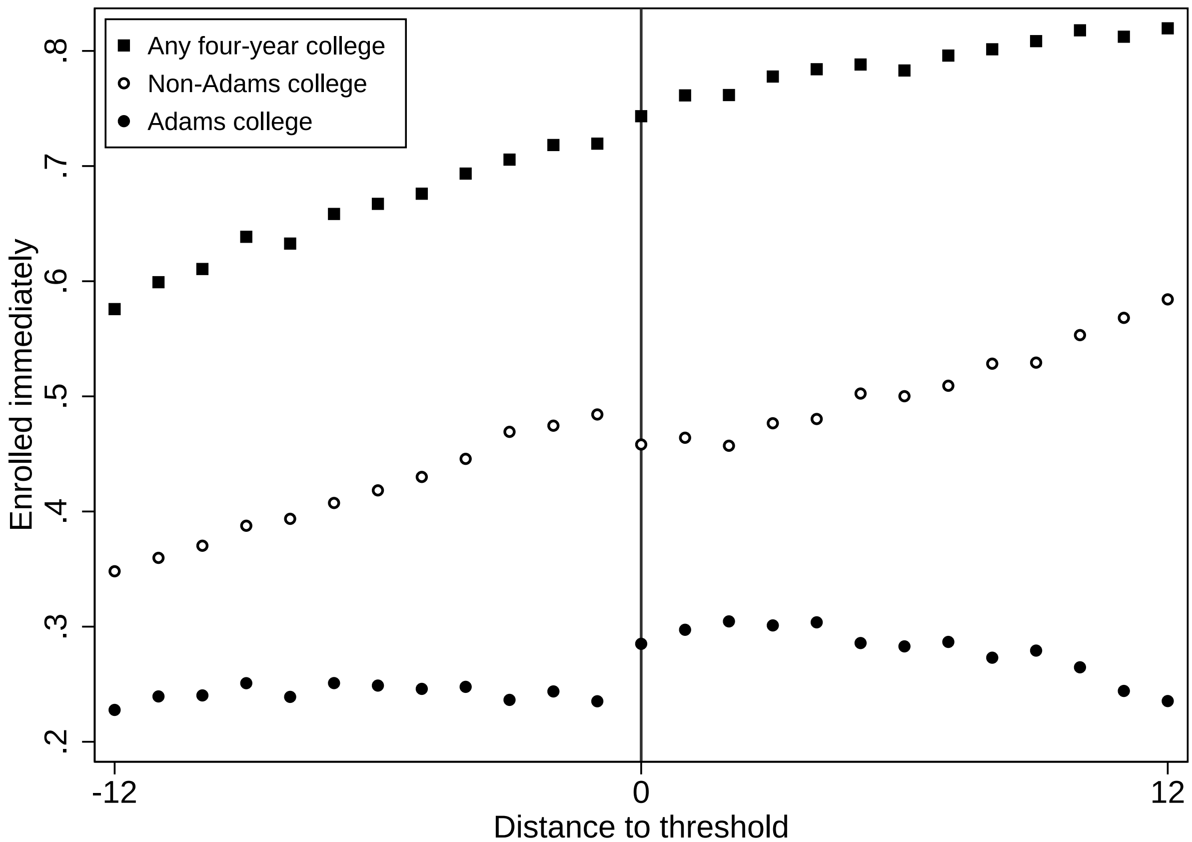
Figure C.10: *Smoothness of Covariates, Classes of 2005-08*





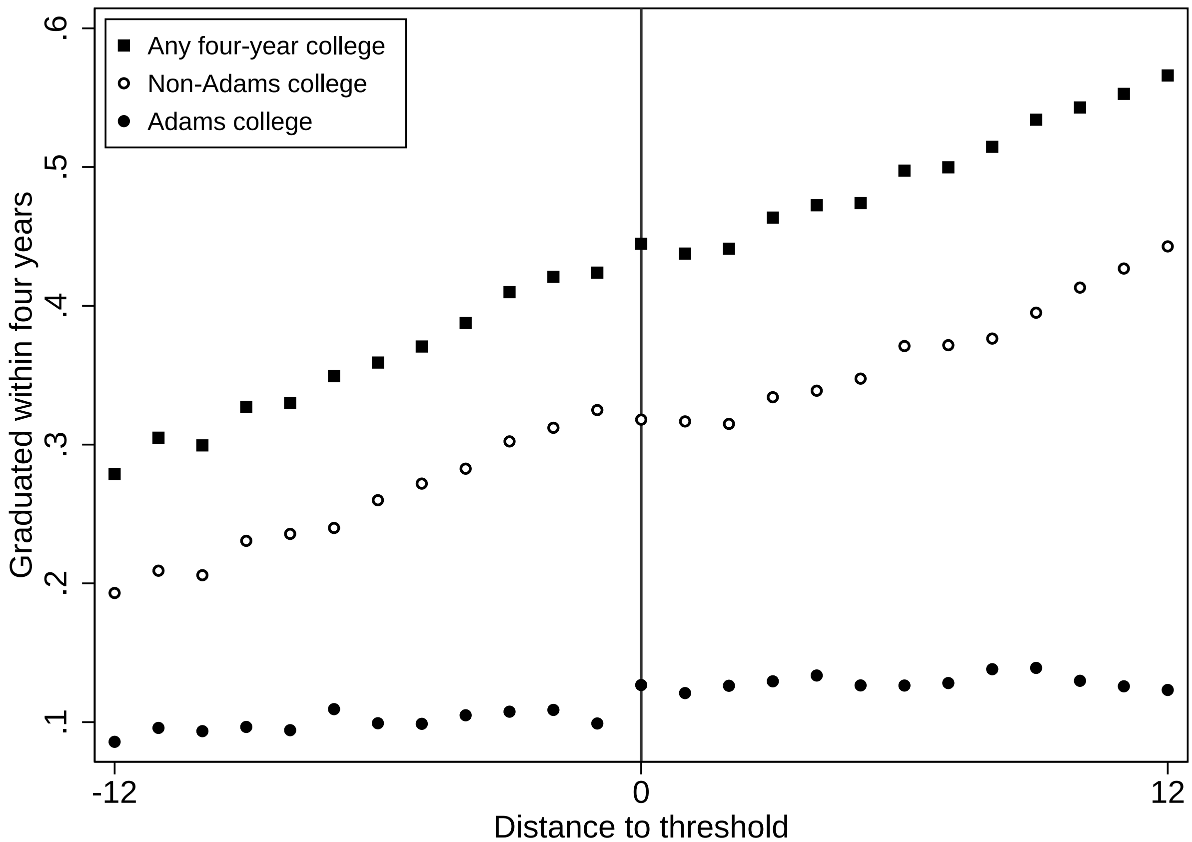
Notes: Each panel shows the mean predicted math and ELA score by each value of the running variable, for the high school classes of 2005-08. Predicted scores are generated by regressing math and ELA scores on demographic characteristics for the class of 2004. The resulting coefficients are then used to generate predictions for subsequent classes.

Figure C.11: *Enrollment at Four-Year Colleges, Classes of 2005-08*



Notes: The above figure shows the fraction of students enrolling in four-year colleges immediately following high school graduation by each value of the running variable, for the high school classes of 2005-08. Adams colleges are Massachusetts public four-year colleges where the Adams Scholarship tuition waiver may be used. Non-Adams colleges are all other four-year colleges, both in-state and out of state. Calculations are based on National Student Clearinghouse data.

Figure C.12: *Graduation from Four-Year Colleges, Classes of 2005-08*



Notes: The above figure shows the fraction of students graduating from four-year colleges within six years of high school graduation by each value of the running variable, for the high school classes of 2005-08. Adams colleges are Massachusetts public four-year colleges where the Adams Scholarship tuition waiver may be used. Non-Adams colleges are all other four-year colleges, both in-state and out of state. Calculations are based on National Student Clearinghouse data.

Table C.1: *College Quality Measures, Selected Institutions*

	2004 MA freshmen (1)	Four-year grad. rate (2)	SAT math score, p75 (3)	Instr. spending (4)	College quality (5)	Net price (6)
(A) Adams colleges						
U. Mass. Amherst	2608	.43	630	9.9	.02	8.6
U. Mass. Dartmouth	1078	.26	580	5.3	-.72	10.7
U. Mass. Lowell	793	.24	610	6.4	-.58	7.6
U. Mass. Boston	349	.12	560	8.8	-1.03	8.3
Bridgewater State	959	.23	560	3.7	-.93	7.7
(B) Other colleges						
Suffolk Univ.	420	.35	550	12.2	-.53	23.3
Univ. of Rhode Island	287	.35	600	7.3	-.38	19.4
Johnson and Wales Univ.	436	.42	590	7.2	-.32	16.9
Univ. of Connecticut	275	.45	650	13.2	.26	18.3
Merrimack College	231	.45	590	7.5	-.22	15.6
Univ. of Vermont	228	.50	630	10.8	.18	18.1
Univ. of New Hampshire	502	.54	620	8.9	.17	19.7
Syracuse Univ.	216	.66	670	16.8	.88	17.7
Boston Univ.	587	.62	690	32.5	1.33	17.2
Tufts Univ.	186	.84	740	29.1	1.96	15.1
Harvard Univ.	124	.86	790	107.8	4.35	12.3

Notes: College characteristics are taken from IPEDS and are measured in the fall of 2004. Instructional spending and net price are measured in thousands of dollars. College quality is the standardized first principal component of each institution's four-year graduation rate, the 75th percentile SAT math score, and instructional expenditures per student, measured as of 2004.

Table C.2: *On-Time Enrollment and Graduation, By Adams College*

	Adams college (1)	Any U. Mass. (2)	U. Mass. Amherst (3)	U. Mass. Dartmouth (4)	U. Mass. Lowell (5)	U. Mass. Boston (6)	Any state college (7)	Bridgewater State Coll. (8)
(A) Classes of 2005-06								
Enrolled immediately	0.069***	0.040***	0.030***	0.009**	-0.000	0.001	0.029***	0.009***
	(0.010)	(0.008)	(0.006)	(0.004)	(0.003)	(0.002)	(0.006)	(0.003)
$\bar{Y}$	0.238	0.151	0.073	0.034	0.031	0.013	0.087	0.024
Graduated within 6 years	0.029***	0.018***	0.015***	0.005	-0.004	0.001	0.011**	0.000
	(0.008)	(0.007)	(0.005)	(0.003)	(0.003)	(0.002)	(0.005)	(0.003)
$\bar{Y}$	0.184	0.119	0.064	0.019	0.023	0.012	0.065	0.026
N	41,190	41,190	41,190	41,190	41,190	41,190	41,190	41,190
(B) Classes of 2005-08								
Enrolled immediately	0.060***	0.036***	0.019***	0.007***	0.003	0.006***	0.024***	0.008***
	(0.007)	(0.005)	(0.004)	(0.003)	(0.002)	(0.002)	(0.004)	(0.002)
$\bar{Y}$	0.235	0.145	0.074	0.034	0.026	0.011	0.090	0.023
Graduated within 4 years	0.018***	0.010***	0.007**	0.001	0.000	0.000	0.009***	0.002
	(0.004)	(0.004)	(0.003)	(0.002)	(0.001)	(0.001)	(0.003)	(0.001)
$\bar{Y}$	0.099	0.067	0.045	0.013	0.007	0.002	0.032	0.009
N	88,152	88,152	88,152	88,152	88,152	88,152	88,152	88,152

Notes: Heteroskedasticity robust standard errors clustered by 12th grade school district are in parentheses (* p<.10 ** p<.05 *** p<.01). Outcomes are enrollment in or graduation from the specific college or set of colleges listed at the top of each column. Each coefficient on aid eligibility is generated by local linear regression with a triangular kernel of bandwidth 12. In panel (A), the sample consists of the high school classes of 2005-06. In panel (B), the sample consists of the high school classes of 2005-08. Listed below each coefficient is the mean of the outcome for students just below the eligibility threshold.

Table C.3: *Robustness Checks*

	Classes of 2005-06			Classes of 2005-08		
	Enrolled immed., Adams college	On campus in year 4, four-year college	Graduated within 6, four-year college	Enrolled immed., Adams college	On campus in year 4, four-year college	Graduated within 4, four-year college
	(1)	(2)	(3)	(4)	(5)	(6)
(A) LLR, no controls						
IK bandwidth	0.067***	-0.022**	-0.023**	0.060***	-0.017**	-0.009
	(0.009)	(0.010)	(0.009)	(0.007)	(0.007)	(0.007)
BW	14.3	10.2	10.3	11.3	9.5	9.4
BW = 6	0.073***	-0.014	-0.013	0.056***	-0.016*	-0.002
	(0.014)	(0.013)	(0.014)	(0.009)	(0.010)	(0.010)
BW = 9	0.072***	-0.020**	-0.019*	0.059***	-0.017**	-0.008
	(0.011)	(0.010)	(0.010)	(0.007)	(0.007)	(0.008)
BW = 12	0.069***	-0.023***	-0.025***	0.060***	-0.017***	-0.010
	(0.010)	(0.009)	(0.009)	(0.007)	(0.006)	(0.007)
BW = 15	0.067***	-0.021**	-0.025***	0.059***	-0.016***	-0.012**
	(0.009)	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)
BW = 18	0.066***	-0.018**	-0.024***	0.059***	-0.015***	-0.012**
	(0.009)	(0.008)	(0.008)	(0.006)	(0.005)	(0.005)
(B) LLR, BW = 12						
District*class FE	0.065***	-0.021**	-0.020**	0.053***	-0.015**	-0.003
	(0.010)	(0.008)	(0.009)	(0.006)	(0.006)	(0.006)
Demographics	0.069***	-0.022**	-0.022**	0.060***	-0.018***	-0.011*
	(0.010)	(0.009)	(0.009)	(0.007)	(0.006)	(0.007)
Donut hole	0.079***	-0.030***	-0.036***	0.071***	-0.025***	-0.020***
	(0.011)	(0.011)	(0.011)	(0.008)	(0.007)	(0.007)
(C) Parametric						
Quadratic	0.063***	-0.028***	-0.029***	0.051***	-0.026***	-0.023***
	(0.009)	(0.008)	(0.008)	(0.006)	(0.006)	(0.006)
Cubic	0.068***	-0.028***	-0.035***	0.061***	-0.019***	-0.014**
	(0.011)	(0.010)	(0.010)	(0.007)	(0.007)	(0.007)
Quartic	0.063***	-0.029**	-0.029***	0.057***	-0.015*	-0.004
	(0.012)	(0.012)	(0.011)	(0.008)	(0.008)	(0.009)

Notes: Heteroskedasticity robust standard errors clustered by 12th grade school district are in parentheses (* p<.10 ** p<.05 *** p<.01). In panel A, each coefficient on aid eligibility is generated by local linear regression with a triangular kernel of the listed bandwidth. Each row in panel B replicates the panel A specification using a bandwidth of 12, with modifications. The first row adds school district by class fixed effects. The second row adds controls for gender, race, low-income status, limited English proficiency and special education status. The third row excludes observations on the eligibility threshold. Panel C fits quadratic, cubic and quadratic functions on either side of the threshold, using the entire sample and a rectangular kernel. The samples consists of the high school classes of 2005-06 and 2005-08.