The Impact of Integration on Academic Achievement: A Study of Boston's Metco Program

Noa Simon

April 20th, 2022

Advisor: Professor Caroline Theoharides

Submitted to the Department of Economics at Amherst College in partial fulfillment of the requirements for the degree of Bachelor of Arts with honors.

Abstract

School integration programs have sparked controversy throughout U.S. history. The Metco Program is no exception. Founded in 1966, Metco sends mainly Black and Hispanic students from Boston to the predominantly White suburban school districts that are nearby. This thesis examines a common criticism of Metco: that district participation negatively impacts overall academic achievement. To evaluate this claim, I implement a Propensity Score Matching (PSM) approach to compare academic achievement in Metco districts to academic achievement in non-Metco districts with similar observable characteristics. Using a panel regression with school year fixed effects, I find that Metco is associated with large and statistically significant increases in overall student achievement at the district- and grade-levels. Controlling for differential trending and school district quality suggests that Metco's seemingly positive effect on achievement may partially be explained by unobservable variables. While the underlying mechanisms of Metco's link to achievement are unclear, I conclude that participation does not negatively impact district-level achievement.

JEL Classification: I21, I24, I29
Keywords: Metco; School Integration; Busing; Educational Inequality; Education

Acknowledgements

Professor Theoharides – I can't thank you enough for your invaluable advice, encouragement, and responses to my many (many) emails throughout this thesis process and the last three years. I am so glad that I was assigned to your intro class - if I hadn't been, I almost certainly never would have even considered writing this thesis.

Mom, Dad, Nate and David – I will never be able to thank you enough…for everything. Dad: enjoy a brief break from your decade-long position as my proofreader/editor-in-chief.

Indie – I couldn't have asked for a better best-friend. The constant support, laughs, and pieces of gum helped me survive this process (and college).

Emily – I am so lucky to have had you as a cheerleader since kindergarten. Should I submit this to the alumni newsletter?

Annie – Thank you for the Stata assistance, but also for your general positivity when I needed it.

Metco Inc. & Department of Elementary and Secondary Education – I am very grateful to Metco administrators Colin Stokes, Glavia Smith, Renee Smith, Patricia Timmons, Dr. Keith Lezama, Latoya Ogunbona, and Tony Laing for taking the time to thoughtfully speak with me about the Metco Program. This thesis also would not have been possible without the help of Scott Kelley and the research team from the Massachusetts Department of Elementary and Secondary Education who answered a variety of data-related questions.

Table of Contents

1 Introduction

Nearly 70 years have passed since the 1954 Supreme Court case *Brown v. Board of Education* ended de jure school segregation, yet Massachusetts' public schools remain racially segregated. 10.4% of Massachusetts public schools have student bodies that are approximately 90% students of color (Scharfenberg 2020). School segregation is particularly pronounced in the Boston area: approximately 76% of students in Boston public schools are Black or Latino while the suburban schools nearby - some of the highest-achieving and wealthiest in the state - are predominantly White (Semuels 2019).

Massachusetts, and Boston in particular, have a long history of racial segregation in public schools. The Boston School Committee refused to integrate schools for nearly two decades after *Brown v. Board.* Furthermore, Boston became known for racism in the 1970s in part due to the city's violent response to court-ordered desegregation (Semuels 2019). Despite this history of resistance, Massachusetts is also home to the Metco program, the longest-running voluntary desegregation program in the U.S.

Founded in 1966, Metco sends predominantly Black and Hispanic students from Boston to the predominantly White suburban school districts nearby. Since its founding, Metco's mission has been to "...expand educational opportunities, increase diversity, and reduce racial isolation…" for students from racially segregated Boston schools and racially isolated suburban schools (METCO, n.d.). While the response to Metco in the participating suburbs historically has been positive, the program has faced significant criticism throughout its existence. In the early years of Metco, criticisms were brazenly racist: parents expressed concerns that Metco students may bring so-called "inner-city diseases" to their districts (Semuels 2019). More recently, other criticisms have included the concern that Metco students may lower average district test scores due to their own performance and by serving as a negative influence in the classroom (Angrist and Lang 2004).

Criticisms of Metco are not unique - integration programs throughout U.S. history have faced staunch criticism. Opposition to school integration can hinder efforts to achieve educational equality due to the link between school segregation and the Black-White racial achievement gap. The economics literature widely acknowledges that school segregation has a negative impact on the outcomes of Black students (see Reardon 2016; Turner 2019; Billings, Deming and Rockoff 2014; Card and Rothstein 2007). As the Brown court indicated, school segregation has an inherent and disproportionately negative psychological impact on Black students (Reardon 2016). Furthermore, majority Black schools have long lacked the resources of majority White schools, an inequality felt by students in Boston, among other U.S. cities, in the decades after the Brown decision (Delmont 2016).

This thesis examines a common criticism of Metco and of integration programs more broadly. The main question I will answer is: What is the effect of district Metco participation on district-level test scores? I answer this question in two steps. First, I implement Propensity Score Matching (PSM) to create a control group of non-Metco districts that are observably similar to Metco districts. Second, I use a panel regression with school year fixed effects to estimate Metco's impact on academic achievement, defined as performance on the statewide English Language Arts (ELA) and math Massachusetts' Comprehensive Assessment System (MCAS) exams administered from fall 2007 through spring 2019.

I find that the majority of estimates of Metco's effect at the grade-level are statistically insignificant. However, the results that are significant imply that Metco has a large, positive effect on district-level average academic achievement. This pattern of results is consistent across two versions of the MCAS exam and is observed when considering the scores of all students and of White students. However, an analysis of a subsample of the data shows a negative, albeit insignificant association between Metco and Black student achievement compared to matched control districts. Controlling for differential trending and school district quality suggests that Metco's seemingly positive effect on achievement may partially be explained by omitted variables. While the factors that drive Metco's link to achievement are unclear, I conclude that participation in the program does not negatively impact achievement.

The structure of this thesis is as follows: I summarize relevant literature, provide Metco background information, and lay out a theoretical framework in sections 2, 3 and 4. Sections 5 and 6 outline my data and empirical strategy respectively. Finally, I summarize and discuss results in sections 7, 8, and 9. All referenced tables and figures can be found in section 11.

2 Previous Literature

There is a significant body of economics literature that explores the effect of school integration on student achievement. As discussed in the introduction, school segregation is linked to unequal outcomes for Black and White students. Conversely, school integration is recognized to improve Black student outcomes (Guryan 2004).

Previous research explores why school integration benefits Black students. One main hypothesis is that integration improves outcomes by increasing the quality of schools that Black students attend (Guryan 2004). Rivkin (2000) offers support for this hypothesis by finding that improvements in school quality are associated with increases in Black student test scores, educational attainment, and future labor market outcomes. Alternatively, school integration may elevate achievement by changing the racial composition of the schools attended by Black students. Research in the '60s offered evidence that racial isolation was detrimental to Black student outcomes (Rivkin 2000). Furthermore, Hanushek, Kain & Rivkin (2009) conclude that schools with higher proportions of Black students have lower math test scores and find that this negative effect of racial composition is approximately two times as large for Black students.

Given the widely acknowledged relationship between race and socioeconomic status in the U.S., integration may alter a school's student body economic composition. Previous studies suggest that this may affect academic achievement: Reardon (2016) finds that Black students' relatively higher level of exposure to low-income classmates is the measure of segregation most highly correlated with the racial achievement gap. Metco's mission of combatting educational inequality through promoting integration and by exposing Boston students to high-quality schools aligns with previous research. Literature on the relationship of student body demographics to achievement is informed by the broader work done on peer effects in education. Peer effects literature suggests that one's achievement is influenced by the achievement of others, though this effect can vary based on the individual student characteristics (Sacerdote 2011). For example, Hoxby (2000) finds that an increase in peer achievement increases individual achievement, and that these effects are larger between students of the same race.

Metco not only aims to expand educational opportunities for students of color, but to improve all student outcomes by increasing diversity. The economics literature on the effects of diversity on student achievement is mixed: Mar (2018) does not find a statistically significant relationship between test scores and racial diversity. In contrast, Tam and Bassett (2004) show that students from diverse high-schools have higher GPAs than students from non-diverse high-schools in their first semester of college at the University of Illinois at Chicago (UIC). Additionally, previous research casts doubt on whether school integration promotes interracial contact: intra-school racial segregation may persist, even if schools integrate (Card and Rothstein 2007; Rivkin 2000).

In addition to the broad literature on integration and peer effects, this thesis is informed by Angrist and Lang's 2004 evaluation of Metco. Their work primarily addresses the claim that Metco students lower the achievement of their non-Metco peers through disruptive behavior and through requiring a disproportionate amount of school resources (Angrist and Lang 2004). In other words, Angrist and Lang (2004) ask whether Metco generates a negative peer effect. Using microdata from Brookline, Massachusetts, the authors find that Metco students do have a significantly negative impact on average school test scores. However, Angrist and Lang suggest that this is not driven by a negative peer effect, but by the fact that Metco students have substantially lower scores than non-Metco students in Brookline. The authors conclude that there is no significant impact of Metco student achievement on non-Metco student achievement. Finally, Angrist and Lang also point that Metco participants demonstrate greater increases in achievement over time than non-Metco students, saying that Metco appears to be beneficial.

I expand Angrist and Lang's investigation in two ways. First, I examine Metco's effect on district-level student achievement across all participating districts, rather than focusing on Metco's effects in Brookline at the school- and classroom-levels. Second, Angrist and Lang measure student achievement using the Iowa Test of Basic Skills (ITBS) test administered to 3rd, 5th, and 7th graders in 1995 and 1996, whereas I rely on Massachusetts' Comprehensive Assessment System exams (MCAS) from school year 2007-'08 through 2018-'19. A more recent analysis of Metco's impact on all districts will offer further insight into the program's effects.

3 Metco Program Background

Metco was founded amidst Massachusetts' reckoning with the *de facto* school segregation that occurred throughout the state, and that was heightened in Boston. In 1965, Massachusetts made school segregation unlawful through the Racial Imbalance Act (RIA). The Act defined a segregated school as a school in which over 50% of the student body was of the same race and 44 schools in Boston met this criteria (Boston Research Center, n.d.). However, the Boston School Committee failed to comply with the RIA and resisted demands to integrate. Furthermore, the Committee failed to provide equal educational opportunities for Black and White students: schools that predominantly served Black students lacked adequate resources yet received significantly less funding than majority White schools (Boston Research Center, n.d.).

While the city was ultimately forced to integrate schools in 1974, the Boston School Committee's initial resistance forced parents to act. In 1965, parents Ellen Jackson and Elizabeth Johnson led a group of 400 Black students to enroll in predominantly White schools with available seats, in a movement called Operation Exodus (METCO, n.d.). Operation Exodus sparked the interest of nearby suburban school districts. Districts subsequently discussed the possibility of enrolling Black Boston students in their predominantly White schools. Metco was founded after months of talks between district representatives, NAACP activists, and the state and federal Departments of Education (Batson & Hayden, 1987). Metco officially began in the 1966-'67 school year with 220 students and 7 partner school districts. By the '70s, the program grew to serve approximately 3,000 students (Semuels 2019).

Since the '70s, the size of the program has remained consistent: Massachusetts' data reveals that in 2021, 3,117 Metco participants attended schools in 37 school districts. The majority of partner districts are located in the greater Boston area and receive Boston students, while a few serve the Springfield school district (Angrist and Lang 2004). The partner districts tend to be high-achieving as well as predominantly White and wealthy (Angrist and Lang 2004). The program is administered by Metco, Inc., a non-profit overseen by the Department of Secondary and Elementary Education (DESE) and funded by a combination of private donors and the Massachusetts government (METCO, n.d.).

The only requirement for Metco eligibility is residence in Boston.[1] Metco students come from a variety of economic backgrounds and are predominantly Black and Hispanic. In the 2019-'20 school year, 67% of Metco students were Black and 24% were Hispanic (METCO, n.d.). Since the 2020-'21 school year, acceptance into the Metco program has been determined by a random lottery (METCO, n.d.; Vaznis 2019). Before then, Metco applicants were accepted on a first-come first-served basis and parents would often sign their children up as infants and toddlers (Vaznis 2019). In conversations with Metco administrators the program was referred to as a "Marginal Seat Program," in which partner districts determine the number of seats available for Metco students largely based on the expected number of non-Metco students in the district in a given grade in each school year. Interviews with Metco administrators revealed that admissions process occurs as follows: Once districts determine the number of available seats, they reach out to Metco, Inc. with their requests. Metco applicants are then referred to districts based on their randomly assigned lottery number and the order in which districts reach out. Families cannot express a preference for a Metco district at any point. The lottery numbers do not give families any type of district preference and are only used to determine the order in which students are referred as districts reach out.

[1] Residence in Boston is specifically required for the Boston-based branch of Metco that is explored in this thesis. Similarly, residence in Springfield is required to participate in the Springfield branch of the program (Springfield Public Schools, n.d.).

Once applicants are referred, the individual districts conduct their own screening processes to decide whether to accept an applicant. Families can only be referred to one district during each application cycle. If a student is not accepted to the district to which they are referred, they can reapply the next year. However, if a student is accepted, parents then choose whether to accept or decline the offer.

Conversations with Metco administrators revealed that screening processes vary by district. Some districts may consider factors such as student attendance rates or whether available extra-curricular align with student interests. Other districts prefer to accept students when they are younger and some may choose not to accept students that would soon need to transition from one school to the next (such as in 8^{th} graders, for example). Despite this variation, kindergarten and 1^{st} grade are typically the years available with the most seats and consequently, many students enter the program at a young age and attend through high school. Within districts, school placement processes vary as well. Some directors expressed that they aim to place Metco students together in the same schools and classrooms, while others did not. Given the above discussion, it is no surprise that Metco personnel described the district-student match process as "decentralized" and as having a "high-level of randomness" (Stokes, Colin. 2021. Interview with author, December 14). Districts receive a grant from the Department of Elementary and Secondary Education (DESE) each year that varies in size based on the number of Metco students in a district. Specifically, DESE allocates funding based on the average number of Metco students enrolled in the district in the last 3 years. This funding scheme somewhat limits the ability of districts to adjust the number of students they accept each year. While funding changes each year, as a frame of reference, in 2019, the average grant given per student was

$4,147, though some districts can receive additional funding as well (Semuels 2019). Funds are used for various purposes, such as for administrative staff and transportation costs (DESE 2017).

The effects of Metco on the participating Boston students are varied. Descriptive analyses suggest that the program improves the academic outcomes of Metco students. Mantil's (2018) analysis suggests that Metco students have higher ELA test scores, graduation rates, and college enrollment rates than their peers in Boston public schools. Furthermore, a conversation with the Brookline Metco administrator revealed that Metco students have performed similarly to their White peers in the district, suggesting that Metco may be effective in closing the racial achievement gap. It is possible that these results are, in part, driven by the positive selection of Metco students: Students that choose to travel a long distance to an academically rigorous school may be more likely to be academically high-achieving. Additionally, parents that choose to enroll their children in Metco may emphasize education at home, a factor that can impact student achievement. However, positive selection is likely not the whole story given that screening processes do not focus on academic achievement and that Metco is a widely publicized program. Despite the academic opportunities afforded by program participation, Metco students face many challenges. Students as young as five endure long commutes that can begin with a bus ride as early as 6 am. Other students report academic difficulties and frequent experiences of racism in their predominantly White school districts (Semuels 2019).

Finally, the effect of Metco on the participating suburban school districts has been debated. As mentioned earlier, while Metco typically enjoys support in the districts in which it operates, it is not without controversy. Parents have expressed concerns that Metco students negatively impact non-Metco students through requiring a disproportionate amount of teacher support in the classroom (Angrist and Lang 2004). Furthermore, while districts receive funding for the Metco students they accept, a common misconception of the program is that it drains districts of resources (Semuels 2019). Others raise the concern that Metco harms Boston Public Schools by drawing high-achieving students to other districts (Angrist and Lang 2004).

4 Theory Framework

There are three broad possibilities for how Metco participation may affect district-level academic achievement: Metco may 1) lower average test scores; 2) raise average test scores; or 3) have no effect on average scores.

A district's participation in Metco will raise average test scores if Metco students perform above the district's average and will lower test scores if they perform below a district's average. If Metco students receive average scores, then Metco will have no effect on average test scores. Consequently, the effect of Metco depends on whether Metco students are from a more positively or negatively selected portion of the distribution of students in the Metco district in which they attend school and the extent to which the school impacts student scores.

We may expect that Metco students are positively selected because they choose to endure lengthy commutes and long school days, and as a result, are likely more motivated than the average student. Additionally, a student's acceptance to Metco implies that administrators believe the student will thrive in an academically rigorous environment. If Metco students are positively selected, their participation may increase average test scores.

Conversely, we may expect that Metco students are negatively selected. One reason we may theorize this is because a lower socioeconomic status has been widely linked to lower levels of academic achievement (Lundstrom 2017; American Psychological Association, n.d.). Metco students likely come from lower-income families than their non-Metco counterparts because Metco districts are some of the wealthiest in the state. In consequence, Metco students may be more likely to receive lower scores than their non-Metco peers. An additional reason we may expect Metco students to be negatively selected is because they arrive from Boston, a district with lower levels of academic achievement compared to Metco districts. My analysis of school year 2018-'19 achievement data reveals that the average English Language Arts (ELA) score in the Boston school district was approximately 17 points lower than the average in Metco districts. Even if Metco students are above-average students in Boston, they may receive lower scores than their non-Metco peers, especially when they begin the program. Furthermore, Metco students may receive lower scores than their non-Metco peers because they face a unique set of challenges: In addition to the taxing commute, Metco students have reported experiences of racism and difficulties adjusting to their new environment (Semuels 2019).

Finally, it is also possible that Metco students have a distribution of scores similar to that of their non-Metco peers. Because many Metco students enter the program in elementary school, and presumably receive the same quality of instruction, we may expect that Metco students will perform similarly to their peers. Students that enter Metco when they are older and need to "catch up" may be able to do so quickly because the schools they enter are well-resourced. If the scores of Metco students drive the effect of Metco participation on average district test scores, we will expect to see that Metco affects Black and Hispanic student achievement, but not White student achievement because Metco students are predominantly Black and Hispanic.

Peer effects theory offers an alternative explanation for how Metco participation may affect district-level test scores. Peer effects theory posits that an individual's peer group affects their individual academic achievement, alongside other determinants of achievement such as family and school resources (Sacerdote 2011; Hanushek and Taylor 1990). A positive peer effect would include a high-achieving student teaching others and a negative peer effect would include a disruptive student distracting others (Sacerdote 2011). As mentioned earlier, some critics claim that Metco generates a negative peer effect due to the teacher attention required by Metco students and due to Metco student misbehavior. However, in their evaluation of Metco, Angrist and Lang (2004) conclude that there is little evidence of peer effects. Furthermore, we may expect peer effects to vary by student race. Hoxby (2000) provides evidence that peer effects are stronger intra-race, suggesting that participation in Metco may only affect the scores of Black and Hispanic students in the district. This theory is consistent with Angrist and Lang's (2004) finding of no significant effect of Metco students on White student scores, but a significant effect on the scores of female Black third-graders.

In summary, Metco students may affect average district achievement due to their own academic achievement or because they influence their non-Metco peers. I hypothesize that districts that participate in Metco will experience decreases in average test scores. I expect that this will occur because Metco students may be more likely to have lower scores than their non-Metco peers.

Finally, while the main focus of this thesis is the effect of Metco on the receiving districts, I want to briefly discuss the possible effects of Metco on the Boston students that participate. The basic education production function models academic achievement as the product of a combination of school resources and family resources (Hanushek and Taylor 1990). Metco sends students to schools and districts that are widely recognized as high-achieving and wealthy. Thus, Metco participation is likely to increase the level of school resources students are exposed to and consequently, positively impact student academic achievement. The notion that Metco may improve student achievement is supported by the fact that Metco students exhibit higher rates of high school graduation and higher test scores than their counterparts in Boston Public Schools (DESE 2018; Mantil 2018).

5 Data Section

My full dataset consists of 33 Boston Metco districts and 292 non-Metco districts in Massachusetts. As I will discuss later, the matched samples include the 33 Boston Metco districts and 21 non-Metco districts in total. I focus on the Boston branch of Metco because of the smaller scope of the Springfield program. Consequently, I remove districts that participate in the Springfield Metco from the sample. My analysis utilizes data from three main sources: The Massachusetts' Department of Elementary and Secondary Education (DESE), the National Center for Education Statistics (NCES), and the U.S. Census Bureau. I include all districts that are included in the U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) data. Because this dataset excludes charter, agricultural, and vocational school districts, these types of districts are excluded from my analysis. Additionally, I exclude four school districts that are not charter nor vocational districts that appear in the DESE data, but not in the Census data. The charter, vocational, and other excluded districts do not appear comparable to Metco districts, and thus, their exclusion is unlikely to affect the control group of non-Metco districts I create and analyze.

5a. Achievement Data

I use DESE's district-level data on the spring Massachusetts Comprehensive Assessment System (MCAS) to measure school district academic achievement from the 2007-‘08 school year to the 2018-‘19 school year. The 2014-‘15 and 2015-‘16 school years are excluded from analysis because some Massachusetts districts implemented an experimental exam during these years.

The MCAS exam is a statewide assessment for students in grades 3 through 10 that evaluates students' understanding of concepts and skills in English language arts (ELA), math, and science. Since 1993, all public-school students have been required to take the MCAS exams for the grades in which they are enrolled (DESE 2020). Students in grades 3 through 8 take the ELA and mathematics exams. Students in grades 5, 8, 9 and 10 also take science exams.

In the 2016-'17 school year, Massachusetts rolled out a new version of the MCAS referred to as the Next-Generation (NextGen) exam. The NextGen exam was implemented for different subjects and grades in different years.[2] The previous version of the exam is referred to as the Legacy exam. In contrast to the Legacy exam, the NextGen exam is intended to be taken online and is described by DESE as offering a "clearer signal of readiness for the next grade level or college and career" than the Legacy exam (DESE 2017). For both the NextGen and Legacy exams, a range of points scored on the exam is attached to four achievement levels.

The four achievement levels on the Legacy exam are advanced, proficient, needs improvement, and warning/failing. The NextGen exam achievement levels are exceeding expectations, meeting expectations, partially meeting expectations, and not meeting expectations (DESE 2017). Discussions with the DESE research team revealed that the two tests are not directly comparable due to the differences in achievement levels and score ranges. As a result, I analyze the effect of Metco on Legacy and NextGen scores separately. Consequently, I analyze the effect of Metco on Legacy achievement from the 2007-'08 school year through the 2013-'14 school year and the effect of Metco on NextGen achievement from 2016-'17 through 2018-'19.

[2] In the 2016-17 school year, the NextGen test was given for the ELA and math exams for grades 3 through 8. However, the grades 5 and 8 science exams and all grade 10 exams remained the Legacy exam. In 2019, the NextGen exam was rolled out for grades 5 and 8 and grades 10 ELA and math, though high school science remained the legacy exam.

For each year the MCAS is administered, DESE reports the number and percent of students that score at each achievement level. Data is reported at the state-level, district-level, and the district-grade level. Achievement is further broken down by 35 student subgroups. Student subgroups include gender, race, and socioeconomic status-related categories. Thus, the broadest category reported is the percentage and number of all students across all grades that score at each achievement level. One example of the most detailed level of data one can look at would be the number and percent of White students and the number and percent of Black students in 3rd grade that score at each achievement level in a given school year.

5b. Achievement Data Limitation

DESE suppresses achievement data for groups of under 10 students. Thus, if there are under 10 students in a given grade-subgroup combination, scores for that combination will be suppressed. Due to this suppression rule, there is a discrepancy in the achievement data reported for Black and White students. For example, in the 2007-‘08 school year, all Metco districts with elementary schools report the achievement of White students in 3rd grade on the Legacy exam. In contrast, only 13 Metco districts report the scores of 3rd grade Black students. While the number of districts that report Black student subgroup scores varies in by school-year and grade-level, this discrepancy limits my ability to consider differential effects of Metco on district-grade level achievement by race. That being said, in the 2016-‘17 through 2018-‘19 sample period, there is significantly more data available for Black student achievement. For example, in the 2016-‘17 school year, 31 of 33 Metco districts report Black student scores for an aggregated measure of students in grades 3 through 8. A significantly higher number of non-Metco districts also report Black scores for grades 3 through 8. As I will later discuss, this increased data availability allows me to consider how Metco’s effect on academic achievement varies by student racial group.

5c. Covariate Data

Data on a variety of covariates are used to assign school districts propensity scores for assignment to Metco. As is further discussed in my empirical strategy, to allow propensity score matching to imitate random assignment, I choose covariates that likely are related both to participation in Metco and district academic achievement. I use covariate data from before the academic achievement sample I analyze to ensure that the variables I use to match districts on are not affected by Metco participation during this period.

The economic variables I include are poverty rate, median family income, and percentage of a district's population with a bachelor's degree. The demographic characteristics I include are a district's urban-centric locale code, the percentage of a district's residents that are White, the percentage of a district's students that are English Language Learners (ELL) and the percentage of a district's students with disabilities. I also include the student-teacher ratio as a measure of school quality.

Several of the covariates are from the NCES Common Core of Data's (CCD) non-fiscal School District Universe Survey from the 2006-'07 school year. I use the CCD data to construct the percentages of a district that are ELL students and students with disabilities. To do so, I divided the number of ELL students and students with disabilities by the number of students in the school district and multiplied by 100. The CCD Data also provides an urban-centric locale code that is a categorical variable consisting of 12 categories characterizing districts as cities, suburbs, or rural areas of different types (for example, code 21 refers to a large suburb). The poverty rate estimates come from the 2005-'06 school year Census Small Area Income and Poverty Estimates (SAIPE) Program. Specifically, I construct this measure by dividing the number of children in poverty by the estimated number of children from ages 5 to 17 that reside, rather than attend school, in the district. The student-teacher ratio measure was retrieved from DESE data from the 2006-‘07 school year.

The percentage of a district that is White and the median family income comes from the 2000 decennial census. These estimates are derived from information on the total population of residents in a district, rather than students that attend school in the district. This ensures that the racial and economic composition estimates of the district are not affected by the district’s participation in Metco.

5d. Summary Statistics Table

Table 1 in section 11 on page 39 summarizes differences between Metco districts and non-Metco districts. Metco districts are higher-income relative to non-Metco districts: the average poverty rate in Metco districts is 0.035 in contrast to 0.091 across all non-Metco districts. The average median family income in Metco districts is nearly $40,000 higher than that of non-Metco districts and the percent of the population with a Bachelor’s degree is approximately 10 percentage points higher in Metco districts than in non-Metco districts. Metco districts and non-Metco districts are similar in terms of their percentages of ELL students and students with disabilities, as well as the population percentage White. Thus, Metco districts are primarily different from non-Metco districts in terms of their economic characteristics. Table 2 on page 40 reveals the general trend that Metco districts have a higher percentage of students scoring at the “advanced” and “proficient” levels than non-Metco districts, and a lower percentage of students scoring at the “needs improvement” and “warning/failing” levels than non-Metco districts. This shows that Metco districts tend to have higher test scores than non-Metco districts.

6 Empirical Strategy

In the ideal experiment, treatment assignment is random and treated and control groups are identical on average. However, school district participation in Metco is not random: Metco districts are high-performing, wealthy, White, and suburbs of Boston. Consequently, sample selection bias threatens estimates of Metco's effect on district-level academic achievement. For example, sample selection bias could potentially result in Metco appearing to increase test scores because the Metco districts are wealthier than non-Metco districts on average, rather than because of Metco's causal impact. Thus, if the issue of sample selection bias is not addressed, estimates of Metco cannot be interpreted as causal.

I use propensity score matching (PSM) to address the likely issue of sample selection bias. Broadly speaking, PSM allows me to compare Metco school districts to observably similar non-Metco school districts. As the name suggests, the PSM approach entails that I create a control group by matching Metco and non-Metco districts together based on their propensity scores. The propensity score is defined as the conditional probability that a school district will participate in Metco. Because propensity scores measure the distribution of a set of characteristics, matching based on the propensity score can create similar-looking treated and control groups (Austin 2011). The intuition of PSM is that if treated and control groups look the same, we can assume that treatment assignment is as good as random. Consequently, the method allows for the imitation of random assignment (Austin 2011; Rosenbaum and Rubin 1983).

The first step of PSM is to estimate the propensity score for each district. As is typical in the PSM literature, I estimate the conditional probability that each district will participate in Metco using the probit function:

$$\Pr(\text{Metco} = 1 | \text{E}_d, \text{D}_d, \text{S}_d, \text{L}_d) = \Phi(\text{B}_0 + \Psi\text{E}_d + \Theta\text{D}_d + \Pi\text{S}_d + \Gamma\text{L}_d) + \varepsilon_d \quad (1)$$

E_d is a vector of economic controls, including district poverty rate, median family income, and the percentage of the district population with a Bachelor's degree. D_d is a vector of demographic controls that include are the percentage of the school district residents that are White, the percentage of ELL students, and the percentage of students with disabilities. S_d is a measure of school quality controls that includes student-teacher ratio. L_d is a vector of location controls that includes an urban-centric locale code from the NCES that describes whether a district is urban, rural, or suburban. This control is in the function as a set of dummy variables.

I include these covariates because they are likely associated both with a district's participation in Metco and the outcome of district academic achievement. It is necessary to include covariates that affect both the probability of treatment assignment and the outcome to allow the Conditional Independence Assumption (CIA) to hold (Smith 2009). The CIA is the assumption that treatment assignment is random, given a set of observed characteristics. Thus, the CIA is the key assumption allowing for PSM to imitate random assignment (Smith 2009).

Economic and demographic characteristics influence the probability a district will participate in Metco, given that Metco districts are predominantly wealthy and White. Furthermore, it is widely documented within the economics literature that the socioeconomic and racial compositions of school districts are associated with academic performance (Reardon 2016; Hanushek, Kain and Rivkin 2009; Card and Rothstein 2007). I also include the urban-centric locale code because proximity to an urban center may affect both academic performance and the likelihood of participation in Metco, given that Metco districts are suburbs of Boston. I include the percentage of ELL students, the percentage of students with disabilities, and the student-teacher ratio because these characteristics likely influence district-level academic achievement. Even if not clearly linked to Metco participation, characteristics that are related to academic achievement may be correlated with Metco assignment and consequently, their inclusion allows me to minimize bias in estimates (Garrido et al. 2014). As mentioned in the data section, the covariates used to generate propensity scores are from school years prior to the academic achievement analysis sample. Thus, later academic achievement should not determine the covariates, reducing the concern regarding reverse causation.

The second step of the PSM approach is to match Metco to non-Metco districts. There are several matching methods, and I use the single nearest neighbor matching (NNM) method. NNM entails that I pair each Metco district to the non-Metco district with the propensity score that has the most similar value. I use NNM with replacement, implying that I can match the same non-Metco district to several Metco districts. I modify NNM by matching districts to the second nearest neighbor in the event that the first nearest neighbor is a district that does not have an overlapping grade range.[3] I matched districts in two steps. First, I matched Metco elementary and middle school divisions, and consequently, all districts included in this sample have students in grades 3 through 8. Second, I matched Metco high school divisions. This sample includes the Metco districts that only have high school divisions and excludes Metco districts that only have elementary and middle school divisions. It was necessary to complete the matching in two steps to ensure that Metco districts with only high schools were not matched with non-Metco districts that did not have high schools at all. Matching in two steps also prevented me from matching Metco districts without high schools to non-Metco districts that did not have elementary and middle schools, but only high schools.

[3] The following two examples are types of modifications that I made to the matching method: First, if the nearest neighbor of a Metco district with only elementary schools was a district with only middle schools, then the second nearest neighbor with an elementary school was chosen. Second, if, for example, the nearest neighbor of a Metco district with grades K-12 was a district with grades K-5, then the second nearest neighbor with grades K-8 was chosen to prevent the exclusion of grades 6,7,8 from analysis and maximize data availability.

An additional matching method I use as a robustness check is nearest neighbor matching with common support imposed. This is distinct from nearest neighbor matching because imposing common support means that I exclude from analysis Metco districts that have propensity scores greater than the maximum propensity scores of the non-Metco districts. Imposing common support leads to the exclusion of 9 Metco districts from the matched sample used both for elementary and high school analysis. The exclusion of the districts changes the average economic and demographic characteristics of the control and treatment groups slightly but does not affect the remaining matches.

The descriptions of the matching methods imply a second condition necessary for propensity score matching termed the "overlap condition": Metco and non-Metco districts must have propensity scores that overlap (Garrido et al. 2014). Figure 1 on page 39 indicates that while the distributions of propensity scores are distinct, there is overlap in the propensity score distributions of Metco and non-Metco districts.

To estimate the treatment effect of Metco, I run a panel regression with school year fixed effects on the matched sample. The equation I estimate is:

$$Y_{dt} = b_0 + b_1(\text{Treat})_t + \delta_t + \varepsilon_{dt} \quad (2)$$

Y_{dt} represents the academic achievement outcome variables for district "d" in school year "t". Treat is a binary variable that is 0 if the district does not participate in Metco and 1 if the district does participate in Metco. δ_t represents school year fixed effects to control for characteristics that may vary over time, but do not vary between districts. My outcome variables are the percentage of students that score at each achievement level for the Legacy and NextGen exams. I consider these outcomes for different student subgroups and grade-levels.

An additional underlying assumption of PSM is that pairing treated and untreated observations based on their propensity scores balances covariates across the treated and control groups. This assumption allows us to behave as though treatment assignment is random. However, it is possible that the propensity scores of the untreated and treated districts are not sufficiently similar to generate balance even in the matched sample. A threat to identification is the potential for a lack of balance within the matched sample of Metco and non-Metco districts. As seen in Table 3 on page 41, significance tests reveal significant differences in the means of economic, demographic, and school quality characteristics between Metco and non-Metco districts.

While the statistically significant differences imply imbalance, the means of the key variables determining Metco participation and academic outcomes, namely median family income, poverty rate, and the percentage White, are similar between the treated and control groups in the matched sample. For example, the mean poverty rate in the matched sample of control districts is 0.039 in contrast to 0.035, a difference that is unlikely to have a practical impact on district academic achievement. Additionally, the difference in the median family income is approximately $12,000; Metco districts have a median family income of $102,047.00 in contrast to $90,183.92. Despite this difference, both the Metco and non-Metco control districts are relatively wealthy, with mean median family incomes higher than the non-Metco district mean of $64,774.90. Furthermore, Metco and non-Metco control districts have nearly identical percentages of the population that are White. Thus, the comparison of the means shows that despite statistically significant differences, the control group produced by matching looks similar to Metco districts. Additionally, a comparison of Tables 1 and 3 show that matching produces a set of control districts that look far more like Metco districts than the average non-Metco district.

Given the similarity of the means, I move forward in my analyses, recognizing that doing so may limit my ability to interpret results as causal and introduce bias into results. I discuss the extent to which I interpret results as causal in sections 8 and 9. Furthermore, I address the covariate imbalances by controlling for baseline characteristics, as I will discuss in section 8.

7 Results

Tables 4 and 5 on pages 42 and 43 report the estimated treatment effect of Metco on ELA and math achievement on the MCAS Legacy exam. The analysis period for grades 3 through 8 is school year 2007-‘08 through 2013-‘14 and for grade 10, through 2017-‘18.[4]

The coefficients are estimates of Metco’s relationship to the percentage of students that score within each of the four achievement levels. Coefficients can be interpreted as a percentage point increase or decrease in the percentage of students that score at a given achievement level. The results reveal a clear pattern: all estimates of Metco’s effect on the percentage of students scoring at the “advanced” level are positive and all estimates of Metco’s effect on the percentage of students scoring at the “needs improvement” and “warning/failing” levels are negative. This trend is seen across estimates of Metco’s effect on the ELA and math achievement of all students and of the White student subgroup. In contrast, the signs on the estimates of Metco’s effect on students scoring at the “proficient” level are mixed. Estimate magnitudes vary by grade-level but are relatively consistent. For the majority of grade-levels, Metco’s effect on the percentage of students at the “advanced” level hovers between 4 and 5, and nearly all estimates of Metco’s

[4] Grades 3 through 8 and grade 10 have different analysis periods because the NextGen MCAS exam was implemented in the 2016-‘17 school year for elementary and middle school, but in 2018-‘19 for high school. Moreover, Legacy and NextGen results are analyzed separately as they are scored based on a different set of standards. Furthermore, the analysis of Metco’s effect on high school scores uses a different control group than that used for Metco’s effect on elementary and middle school scores since, as discussed earlier, there are two sets of matches to account for different grade-ranges offered by districts.

effect on the percentage of students at the "Needs Improvement" and "Warning/Failing" levels lie between 0 and -2. Estimates of Metco's effect on math achievement tend to have slightly larger magnitudes, though a similar level of consistency is observed.

A clear trend is also observed in terms of the statistical significance of estimates. The majority of estimates are statistically insignificant. However, there are exceptions to this trend. First, as seen in Tables 4 and 5, Metco has a statistically significant positive effect on the percentage of all students, in all grades and on the percentage of White students, in all grades that score at the "advanced" level on both the ELA and math exams. The causal interpretation of the estimates of Metco's effect on ELA achievement are that Metco increases the percentage of all students, in all grades that score in the "advanced" category by 5.25 percentage points and the percentage of White students, in all grades that score in the "advanced" category by 6.18 percentage points. In other words, Metco has a statistically significant positive effect on district-level achievement. These estimates imply large effects. On the ELA exam, the average percent of students in the control districts that score at the "advanced" level is 28.20% and the average percent of "advanced" White students is 27.27%. Thus, at their respective means, Metco increases the percent of all students that score at the "advanced" level by approximately 18.61% and the percent of White students in all grades that score at the "advanced" level by 22.66%. Metco's positive effect on the percentage of all students that score at the "advanced" level on the math exam are similarly large. At the control district average of 36.83%, Metco is associated with a 6.99 (18.99%) percentage point increase in the percent of all students that score at the "advanced" level. At the control district average of White students of 36.45%, Metco's positive effect on the percentage of "advanced" White students implies a 7.846 (21.52%) percentage point increase in the percent of "advanced" students.

Large and significant effects at the grade-level are observed as well: for example, at the control district average of 16.91%, Metco increases the percentage of White 4^{th} grade students scoring at the "advanced" level on the ELA exam by 5.11 (30.24%) percentage points. On both the math and ELA exams, a few estimates of Metco's effect on the percentage of students that score at the "needs improvement" and "warning/failing" levels are significant as well. The effects are negative and of similarly large magnitudes: for example, Metco is associated with a 0.80 percentage point decrease in the percent of White students in all grades that score at the "needs improvement" level on the Legacy exam. At the control district average of 3.07%, this implies a 26.05% decrease. Overall, the results imply that Metco increases the percentage of all students, and White students, that achieve the "advanced" level and conversely, decreases the percentage of students achieving the "needs improvement" and "warning/failing" levels on the Legacy ELA and math exams.

Tables 6 and 7 on pages 44 and 45 report the effect of Metco on NextGen ELA and math exam achievement. The period of analysis here is school year 2016-'17 through 2018-'19. Even though the two tests are not directly comparable, the pattern of Metco's effects on NextGen exam achievement and Legacy exam achievement are similar: nearly all estimates of Metco's effect on the percentage of White and all students scoring at the "exceeds expectations" level across both the math and ELA exams are positive, while estimates on the percentage of students scoring at the "partially meeting expectations" and "not meeting expectations" levels are nearly all negative. The magnitudes of Metco's effect at the grade-level on both the ELAand math NextGen exams are relatively consistent with one-another and with Legacy exam estimates.

In addition, as was observed in the Legacy exam analysis, the majority of estimates of Metco's effect on NextGen achievement are statistically insignificant. However, exceptions to this are estimates of Metco's effect on the percentage of all students and White students that score at the "exceeds expectations" level on the ELA test in the 4th and 5th grade. The ELA NextGen effects are slightly larger than the effect of Metco on the Legacy exam. For example, Metco is associated with a 5.41 percentage point increase in all 4th graders that score at the "exceed expectations" level. At the mean of 13.45%, this implies that Metco increases the percentage of students that score at this level by 40.20%. Furthermore, Metco has a statistically significant negative effect on the percentage of students "partially meeting expectations" in 4th and 5th grade, and "not meeting expectations" in 7th grade. Estimates of Metco's effect on NextGen achievement mirrors the pattern displayed in the Legacy exam results: Metco is associated with increases in the percentage of students that meet and exceed expectations and decreases in the percentage of students that partially and do not meet expectations.

While the majority of estimates are statistically insignificant, the key implication of the results is that Metco is associated with greater student achievement. Initially, such large effects may appear implausible. However, the translation of the percentage point increases to a number of students show how the results may be realistic. For example, the average number of White 4th grade students scoring at the "advanced" level on the ELA Legacy exam in the control group was approximately 32 students. The reported 30.24% increase then would amount to an increase of approximately 10 White 4th grade students.

The assessment of the plausibility of the results depends on the portion of the distribution from which we think students move. It would not be surprising if, as discussed above, 10 4th grade students earned a few more points and thus, moved from the "proficient" to the "advanced" levels. However, we may be more skeptical of the results if Metco prompted students to move from the "warning/failing" level to the "advanced" level. It is difficult to discern the source of the effect of Metco on the achievement of White 4th graders because the positive coefficient on the "proficient" achievement level is insignificant, and coefficients on the "needs improvement" and "warning/failing" levels are both negative. The statistically significant positive effect of Metco on the percentage of White 5th grade students that score at the "exceed expectations" level is accompanied by a statistically significant and large decrease in the percentage of students that score at the "partially meet expectations" level, suggesting that perhaps students move from this portion of the distribution.

As seen in Tables 6 and 7, I compare the scores of all students, White students, and Black students in an aggregate measure of 3rd through 8th grade on the NextGen exam using both a full sample and a subsample modified to accommodate limitations on Black student achievement data. The subsample was created by removing district-year observations in which the scores of all students and White students were available, but not Black students. I changed four districts in the control group to perform this analysis because the initial matches lacked sufficient data. To do so, I picked the districts with the next closest propensity score as new matches. The same pattern of results is observed for all students and White students as is seen in Tables 4 and 5: the coefficients on the "exceeds expectations" and "meets expectations" are positive, whereas the coefficients on the "partially meeting expectations" and "not meeting expectations" are negative. This pattern is consistent across the all student and White students groups, as well as across the ELA and math exams. Additionally, Metco appears to be associated with a statistically significant 30% increase in the percentage of all students in grades 3 through 8 that score at the "exceeds expectations" level. The results for Black students diverge from the otherwise consistent trend: on the ELA exam, Metco is associated with slight increases in the percent of Black students in 3rd through 8th grade that score at the "exceed expectations" and the "partially meeting expectations" levels, and slight decreases in the percent scoring at the "meets expectations" and "partially meeting expectations" categories. In contrast, on the math exam, Metco is associated with decreases in the percent of Black students that score at the "exceeds expectations" and "meets expectations" levels.

There are two main questions raised by the results. First: the seemingly positive effect of Metco raises the question of whether a district's participation in Metco increases the achievement of all students, and White students specifically, or whether there are other unobserved factors that drive these results. Second: why does the program appear to be associated with negative outcomes for Black students. While the estimates of Metco's effect on Black student scores are not statistically significant, this pattern is concerning as it creates the appearance that the program may harm the students it intends to benefit.[5]

8 Robustness Checks

The robustness checks shed light on the extent to which results should be interpreted as causal. Table 8 on page 46 shows estimates of Metco's effect on the scores of all students in all grades when robustness checks are incorporated. I ran checks on various grade-levels and student subgroups but found a consistent pattern and thus, only report district-level estimates.

8a Checks 1 & 1a: Baseline Covariates & Differential Trending

While matching created a control group of non-Metco districts that looked like Metco districts, imbalance in observable characteristics remained after matching was performed. For example, the Metco districts have a median family income that is $12,000 greater than that of the control districts. Such differences may contribute to the association between Metco and positive

---

[5] I must note that I was only able to consider Black student scores on the NextGen exam in the aggregate 3rd through 8th grade due to a lack of data.

outcomes. To address this potential source of bias, I include the pre-analysis period covariates, including median family income, the percentage of the district population with a Bachelor's degree, and the percentage White, as controls in the second-stage panel regression.

An additional possible source of bias is that Metco districts may be trending differentially over time compared to non-Metco districts. This threatens identification because it implies that results may capture differences in how districts change over time, rather than capture the effect of Metco participation. For example, it is possible that Metco districts were affected by the 2008 financial crisis differently than non-Metco districts due to their economic characteristics. Thus, differences in academic performance could be driven by the differential impact of 2008 on districts, for example, rather than by Metco participation. I address the potential for differential trending by interacting school year fixed effects with the pre-analysis period covariates as well.

Both checks resulted in magnitudes of district-level estimates that are slightly smaller than those of the main results reported in Tables 4 and 5. While the effect of robustness checks on the grade-level estimates are not reported in Table 8, I want to note that I also saw this trend when I incorporated this check into the grade-level analysis. Despite differences in estimate magnitudes, the overall pattern remains the same: Metco is associated with a greater percentage of students above proficient and lower percentage of students below proficient.

In terms of statistical significance, the checks performed resulted in district-level estimates that were significant in the main results becoming statistically insignificant. Likewise, I found that the majority of grade-level estimates were insignificant. However, the few estimates that remained significant confirmed that Metco participation is associated with increases in the percentage of students that score above proficient and decreases in the percentage of students that score below proficient.

The fact that the overall pattern of the main results is maintained when these checks are performed mitigates the concern that Metco's association with positive outcomes is entirely driven by imbalances in observable characteristics, such as income, or by differential trending. However, that estimate magnitudes are reduced and that estimates became insignificant suggests that differences between Metco and non-Metco districts, and differential trending may be factors underlying the relationship between Metco and positive student outcomes.

8b Check 2: Imposition of Common Support

A potential threat to the validity of the main results is that I do not impose common support. In other words, I do not exclude the Metco districts with propensity scores that are higher than the maximum propensity score of the non-Metco districts from my matched sample. I initially do not impose common support because the exclusion of the Metco districts changes the treatment effect that I am estimating from the effect of Metco participation on all districts to the effect of participation on some districts (Smith 2009). However, given that the propensity score captures the balance of a set of characteristics, a lack of common support is an issue because it implies that the control group is lacking a "good" match that is observably similar to the treated unit (Smith 2009). The lack of a "good" match threatens the assumption that treatment assignment is as good as random because the treatment and control groups are observably similar. In addition, the majority of Metco districts outside the region of common support have median family incomes that are well above the average in both Metco and control districts. For example, the district Weston which is excluded when I impose common support, has a median family income that is approximately $80,000 higher than the Metco district average. Given the link between income and academic achievement, we may be concerned that the seemingly positive effect of Metco is driven by the higher-incomes of such districts. For these reasons, the inclusion of districts that appear to lack a "good" match may threaten the assumption that treatment assignment is random.

To test this threat, I ran the results again while excluding the 9 Metco districts that are outside of the region of common support. When I impose common support, I find a similar pattern of results. As reported in Table 8, Metco is again associated with district-level increases in academic achievement, though the magnitudes of estimates are slightly reduced relative to the main results. Furthermore, the district-level estimates again become statistically insignificant. While I only report district-level estimates in Table 8, I found that imposing common support also tended to reduce the magnitudes of grade-level estimates and that the majority of estimates significant in the main results became statistically insignificant. Overall, we see that when common support is imposed, the pattern of results is consistent with that of the main results: Metco remains linked to increases in student achievement. However, the fact that the majority of previously significant district- and grade-level estimates become insignificant suggests that perhaps outlying Metco districts play a role in the program's apparent link to positive outcomes in the district. Thus, this check supports the notion that Metco is associated with increases in academic achievement, though calls into question a causal interpretation of results.

8c Check 3: Pre-Period Academic Achievement Check

A primary concern is that the positive effect of Metco may be driven by the high-quality academics of the Metco districts, rather than the effect of Metco itself. I attempt to control for this in my probit specification by using student-teacher ratio as a measure of school quality. However, prior research suggests that student-teacher ratio is an inadequate measure of quality (Rivkin 2000). Furthermore, Table 1 shows that student-teacher ratio is similar for most Massachusetts' districts, while district school quality presumably varies, suggesting that this measure may be inadequate. To strengthen my control for school quality, I add an additional control to my panel regression: 2005 academic achievement. I measure 2005 academic achievement as the percentage of students that score in the advanced category on the ELA exam in 3rd grade for all Metco districts with elementary and middle schools, 7th grade for Metco districts with only middle school divisions, and 10th grade for Metco districts with only high school divisions.

When I include this additional measure of school quality, the pattern of the main results is largely the same. District-level estimates of Metco's effect on the percent of students scoring at the "advanced" level are positive, while estimates of Metco's effect on the percent of students scoring below proficient are negative. When I performed this check on grade-level estimates, I found that nearly all estimates of Metco's effect on the percent of students scoring at the "advanced" and "exceeds expectations" categories are positive and the majority of the coefficients on the achievement levels below "proficient" and "meets expectations" are also negative. However, as seen when the other two robustness checks are performed, the magnitudes on the district-level are somewhat reduced and estimates become statistically insignificant. These results suggest that perhaps school district quality may partially drive the relationship between Metco participation and increases in student achievement. That being said, even once academic achievement is controlled for, the pattern of Metco's effect is the same. Overall, all three robustness checks result in district-level estimates that are reduced and statistically insignificant, but Metco's broad association with positive outcomes remains.

9 Discussion

School integration has been hotly debated topic in American public life for more than 70 years. While integration is widely recognized to benefit students of color, there is debate regarding its' effects on the district-level achievement of school districts populated predominantly by White students. Opponents of Metco have raised the concern that the program negatively impacts district-level academic achievement and the outcomes of non-Metco students in the partner districts. My analysis of the Metco program shows that this perception is false. An analysis of two versions of the MCAS, multiple student groups, and several grade-levels reveals a consistent link between Metco and positive student outcomes on both ELA and math exams: Metco is associated with increases in the percentage of students that score at the "advanced" level and the "exceed expectations" levels on the old and new versions of the MCAS, respectively. Furthermore, participation is associated with decreases in the percentage of students that score below the "proficient" and "meets expectations" levels. While many estimates are statistically insignificant, the estimates that are significant at the grade- and district-level suggest Metco has a large, positive effect on student achievement. When three robustness checks are performed, estimates of Metco's district-level effects are reduced in magnitude and become statistically insignificant, though program participation remains broadly associated with positive student outcomes.

The lack of a negative effect of Metco participation casts doubt on critics' claims that Metco students lower average district achievement because their performance is significantly lower than that of their non-Metco peers, as is suggested by program critics (Angrist and Lang 2004). Furthermore, this lack of a negative effect, and in some cases, a positive effect of Metco participation appears to contradict Angrist and Lang's (2004) finding that increases in the number of Metco students at the grade- and school-level decreases test scores in Brookline, a result the authors attribute to the lower scores of Metco students relative to non-Metco Brookline students. A likely explanation for why my results differ from that of Angrist and Lang is that I consider the effect of Metco participation at the district-level, whereas Angrist and Lang (2004) consider variations in the number of Metco students at school- and grade-level. It is possible that a district-level analysis does not capture the program's effect on test scores because Metco students account for a small percentage of total district enrollment: the size of the program has varied over time, but, as of 2020, the median percentage of total district enrollment comprised by Metco students comprised was 2.28%. Given this small percentage, it is likely that Metco student scores would be absorbed by the average, regardless of whether Metco students performed above or below the average of their non-Metco peers.

An additional explanation for the lack of a negative effect of program participation is that Metco students may not receive lower scores than their non-Metco peers. There are two main reasons why we may not expect Metco students to score lower than their peers. First, while the average academic achievement in Boston is lower than that in the Metco districts, that Metco students are likely positively selected relative to Boston students may result in Metco students performing similarly to their non-Metco peers. Furthermore, given that many Metco students enter the program in kindergarten and 1st grade, it is plausible that their early instruction in the Metco districts closes any achievement gap we may expect to see between Metco and non-Metco students. An additional possibility for why my results differ from that of Angrist and Lang is that Metco student achievement has changed over time. Angrist and Lang (2004) use data from before 2000, whereas I consider scores from 2007 through 2019. A conversation with Brookline's Metco director revealed that over the last five years or so, Metco students have performed similarly to their White non-Metco peers, and in fact have received higher scores than non-Metco students of color. While this was only mentioned by the Brookline director, perhaps this trend has occurred in other districts. Differences between my results and those of Angrist and Lang (2004) may result from changes in Metco student achievement over time.

Metco's association with positive student outcomes also calls into question the criticism that Metco students lower test scores by prompting a negative peer effect. Given that White students comprise only 1% of Metco, we can assume that the vast majority of White students in Metco districts do not participate in Metco (METCO, n.d.). Consequently, the finding that Metco has a statistically significant positive effect on the scores of White students eliminates the concern of a negative peer effect. The seeming lack of a negative peer effect is consistent with Angrist and Lang's (2004) finding that Metco has no significant effect on the scores of White non-Metco students.

An exception to Metco's consistent link with positive student achievement is that the program is associated with small decreases in the percentage of all Black students in grades 3 through 8 that score at the "exceeds expectations" and "meets expectations" levels on the math exam. This suggests that Black students in non-Metco districts receive higher test scores than Black students in Metco districts. However, the results are not statistically significant and a lack of data availability prevents my consideration of a broader pattern. In consequence, that estimates of Metco's effects are insignificant further offers support for the notion that, even if Metco does not necessarily improve district-level achievement, at the very least, participation in the program does not have a significant negative impact at the district-level.

The essential question is whether the results should be interpreted as Metco's causal effect, or the result of differences between Metco and non-Metco districts. There are at least two explanations for why Metco participation may positively affect test scores. First, as was mentioned above, it is possible that Metco students receive above average test scores in the district. The potentially higher scores of Metco students might be associated with increasing district-level test scores. Second, as was discussed in the literature review, Hoxby (2000) provides evidence of intra-race peer effects. If Black Metco students do receive higher scores than their Black non-Metco peers, as suggested by the Brookline director, perhaps they prompt a positive peer effect, which may result in higher test scores on average in the district.

The results of the robustness checks suggest that unobservable variables may bias the estimates. All checks reduce estimate magnitudes and result in district-level estimates becoming statistically insignificant, though the general pattern remains the same. The school academic quality measure results in the largest district-level estimate magnitude reductions, suggesting that perhaps the academic quality of Metco districts contribute to results. This is no surprise: Metco districts are widely recognized as high-achieving and for decades have chosen to participate in a program designed to give Boston students access to a higher quality of education. Thus, it is possible that the seemingly positive effect of Metco on test scores is the result of the academic quality of the districts, rather than the causal effect of Metco itself.

Limitations of propensity score matching to create an identical control group may also impact results. While propensity score matching allowed me to create a control group of districts that looked much more like Metco districts than the average district in Massachusetts, there were statistically significant differences in the included covariates. For example, as discussed earlier, the average median family income in Metco districts is approximately $12,000 more than that of the average median family income in the non-Metco control districts in the matched sample. While all districts in the matched sample are wealthier than the Massachusetts average, the remaining economic differences may be correlated with district academic quality, and student academic achievement. However, when I control for baseline characteristics, including income level, Metco's association with a greater level of achievement is preserved.

It is also possible that differences in socioeconomic status are a cause of the seemingly negative effect Metco has on Black student scores. Black students in the Metco districts are likely to be Metco students whereas Black students in non-Metco districts are likely residents of those districts. Metco students, who are from Boston, may come from less wealthy families than students who are residents of the comparatively wealthy Metco control districts. This potential difference in socioeconomic status may drive Metco's seemingly negative effect on Black student scores. Finally, given that Metco has existed since 1966 in the majority of the districts that participate today, it is possible that the program has changed unobservable characteristics in Metco districts, such as political attitudes for example, that render them fundamentally distinct from other districts in Massachusetts. If Metco and non-Metco districts are distinct based on unobservable characteristics, using propensity score matching may not have sufficiently allowed me to create a control group that is identical to Metco districts. A limitation of this analysis is that I do not observe the academic achievement in the districts before and after they begin to participate in Metco due to the length of the program's existence.

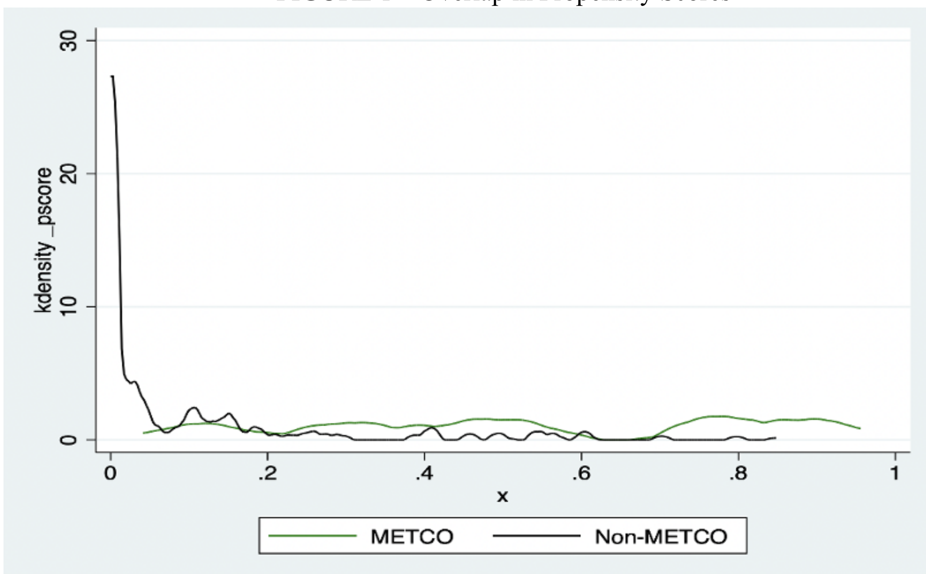
While I hesitate to suggest that a district's participation in Metco increases student achievement, the results do call into question whether Metco participation has a negative impact on achievement. If the program's seemingly positive effect results from the academic quality of the Metco districts, then we at least see there is little evidence that participation in the program reduces the academic advantage beyond non-participating districts.

10 Conclusion

School segregation remains a central issue in Massachusetts' public education. While wealthy suburban school districts have seen increases in their student diversity, they remain predominantly White and Asian (Scharfenberg 2020). Meanwhile, the number of schools in the state with student bodies that are over 90% nonwhite have increased in the past decade (Scharfenberg 2020). School segregation is an obstacle to educational equality: there is a widely documented link between school integration and improved student outcomes for students of color. In consequence, efforts to hinder school integration also hinder efforts to combat educational inequality. Historically, a source of opposition to school integration efforts and specifically to Metco has been the perception that such programs negatively impact academic achievement in the White districts that integrate. This thesis suggests that there is little evidence that Metco negatively impacts district-level academic achievement. On the contrary, Metco participation is consistently associated with positive student outcomes. Even if Metco participation itself does not drive the program's apparent link to positive outcomes, at the very least the results cast doubt on the notion that Metco negatively impacts district academic achievement. If the positive effect of Metco is because participating districts are far better than other Massachusetts districts, then this thesis shows that, at a minimum, Metco participation does not diminish district academic excellence. By calling into question common criticisms of Metco, these results serve to bolster the case for continuing Metco, and similar integration programs, that are crucial to promoting educational equality.

11 Figures and Tables

FIGURE 1 – Overlap in Propensity Scores



NOTES: This shows the distribution of the propensity scores assigned to Metco and non-Metco districts in the full sample. This was generated using the Stata psmatch2 program (Leuven and Sianesi, 2013).

TABLE 1 – School District Summary Characteristics

	Non-Metco Districts		Metco Districts	
	Mean	Sd	Mean	Sd
Bachelor's Degree (%)	19.320	0.608	30.130	4.539
Poverty Rate	0.091	0.608	0.035	0.011
Median Family Income ($)	64,774.90	17114.20	102,047.00	23,599.13
English Language Learners (%)	0.034	0.054	0.017	0.018
Students with Disabilities (%)	0.158	0.032	0.157	0.026
White (%)	91.390	9.560	92.713	4.267
Student-Teacher Ratio (to 1)	13.461	1.822	13.174	1.425
Fraction Suburban	0.661		0.901	
Number of Observations	228,844		34,077	
Number of Districts	292		33	

NOTES: For each district in the sample, there is a unit of observation for each grade-year-student subgroup combination. Statistics on English Language Learners (%), Students with Disabilities (%), Fraction Suburban, Poverty Rate, and Student-Teacher ratio are from school year 2006-'07 data. Statistics on Bachelor's Degree (%), Median Family Income ($), and White (%) are from data published in 2000. English Language Learners (%) and Students with Disabilities (%) are school district student characteristics while the remainder of the variables are characteristics of the school district's total population.

TABLE 2 – School District Academic Achievement Summary

	Non-Metco Districts		Metco Districts	
	Mean	Sd	Mean	Sd
Legacy Exam				
ELA				
% Advanced	14.267	14.505	25.465	18.666
% Proficient	48.922	16.276	51.548	13.232
% Needs Improvement	26.979	15.617	18.165	14.378
% Warning/Failing	9.857	11.158	4.849	5.724
Math				
% Advanced	21.918	17.739	35.542	21.713
% Proficient	30.921	11.851	32.694	11.144
% Needs Improvement	29.237	12.976	22.051	13.697
% Warning/Failing	17.943	16.043	9.736	10.092
NextGen Exam				
ELA				
% Exceeds Expectations	6.196	7.177	13.032	10.910
% Meets Expectations	37.358	16.123	47.043	13.559
% Partially Meeting Expectations	43.363	14.009	33.329	15.907
% Not Meeting Expectations	13.21	12.574	6.614	7.080
Math				
% Exceeds Expectations	5.850	7.166	14.237	12.348
% Meets Expectations	35.238	16.41	45.176	13.737
% Partially Meeting Expectations	44.719	14.01	33.095	16.075
% Not Meeting Expectations	14.481	12.822	7.509	7.492
Number of Observations				
Legacy Exam				
ELA	163,122		20,023	
Math	163,269		20,031	
NextGen Exam				
ELA	75,510		9,269	
Math	75,499		9,271	

NOTES: There is a unit of observation for each grade-student subgroup-year combination for each district in the sample. Summary statistics are of average achievement data from the 2007-‘08 to the 2018-‘19 school years. The sample period excludes the 2014-‘15 school years and 2015-‘16 school years due to the implementation of the experimental PARCC exam in some Metco districts.

TABLE 3 – Covariate Statistical Significance Tests

Covariate	Control Districts	Metco Districts	Difference
Poverty Rate	0.039	0.035	0.005***
Median Family Income ($)	90,581.67	102,047.00	11465.37***
Student Teacher Ratio (to 1)	13.345	13.149	0.196***
English Language Learners (%)	2.314	1.804	0.510***
Students with Disabilities (%)	15.420	15.783	0.363*
Bachelor's Degree (%)	30.061	29.997	0.064
White (%)	92.954	92.436	0.520**
Number of Observations	862	2,086	
Number of Districts	14	33	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

NOTES: Statistical significance tests reported here are performed on the matched sample used for grades 3 through 8 analysis. There is a unit of observation for each grade-student subgroup combination for each district in the sample. As discussed in the empirical strategy, the control districts in this matched sample differ from the control districts used for the grade 10 analysis. However, when I performed statistical tests on the matched sample used for grade 10 analysis, I found a similar pattern of results.

TABLE 4 – Effect of Metco on Legacy ELA Achievement

	Number of Observations	% Advanced	% Proficient	% Needs Improvement	% Warning/Failing
All Students					
All Grades	282	5.250*	-2.131	-2.349	-1.000
		(2.567)	(1.477)	(1.272)	(0.515)
Grade 3	308	1.350	1.014	-1.235	-1.204
		(2.031)	(1.335)	(1.760)	(0.765)
Grade 4	307	5.114*	0.322	-3.716	-1.648
		(2.337)	(1.151)	(1.970)	(0.905)
Grade 5	308	4.796*	-0.729	-2.949	-1.081
		(2.349)	(1.190)	(1.634)	(0.602)
Grade 6	300	3.677	-1.514	-1.538	-0.607
		(2.118)	(1.324)	(1.082)	(0.566)
Grade 7	301	3.915	-1.679	-1.830	-0.325
		(2.621)	(1.862)	(1.148)	(0.417)
Grade 8	300	3.720	-1.688	-1.521	-0.500
		(2.807)	(2.048)	(0.948)	(0.340)
Grade 10	360	7.511	-5.422	-1.554	-0.538
		(3.810)	(2.832)	(0.960)	(0.392)
White Students					
All Grades	282	6.181*	-2.655	-2.884**	-0.799
		(2.525)	(1.519)	(1.062)	(0.330)
Grade 3	308	2.065	0.969	-2.217	-0.959*
		(1.798)	(1.002)	(1.398)	(0.427)
Grade 4	307	5.246*	0.809	-4.572*	-1.519*
		(2.233)	(0.911)	(1.799)	(0.650)
Grade 5	308	5.210*	-0.873	-3.509*	-0.942*
		(2.170)	(1.099)	(1.393)	(0.426)
Grade 6	300	4.284*	-1.563	-2.303*	-0.484
		(1.952)	(1.235)	(0.896)	(0.366)
Grade 7	301	4.472	-2.014	-2.357*	-0.156
		(2.497)	(1.851)	(0.963)	(0.290)
Grade 8	301	4.322	-2.156	-1.686*	-0.428
		(2.698)	(2.037)	(0.828)	(0.257)
Grade 10	360	7.438*	-6.126*	-1.110	-0.282
		(3.422)	(2.724)	(0.680)	(0.287)
Control District Average	806	28.200	54.315	14.020	3.447
		(15.404)	(10.814)	(8.577)	(2.780)

Standard errors in parentheses
*p<0.05, ** p<0.01, *** p < 0.001

NOTES: There is a unit of observation for each district-grade-year combination. The academic achievement analysis period for grades 3 through 8 is from the 2007-'08 school year to the 2013-'14 school year. The grade 10 academic achievement analysis is from 2007-'08 through 2017-'18 because in the NextGen test wasn't implemented for grade 10 until the 2018-'19 school year. Control district average refers to the average achievement of all students in the matched sample used for grades 3 through 8 analysis across all years of the sample period. The sample period excludes the 2014-'15 and 2015-'16 school years due to the implementation of an experimental exam in these years. The full specification includes school year fixed effects and robust standard errors clustered at the district-level.

TABLE 5 – Effect of Metco on Legacy Math Achievement

	Number of Observations	% Advanced	% Proficient	% Needs Improvement	% Warning/Failing
All Students					
All Grades	282	6.995*	-1.548	-3.752*	-1.578
		(3.086)	(1.133)	(1.703)	(0.978)
Grade 3	308	2.971	-0.491	-1.194	-1.337
		(2.996)	(1.090)	(1.701)	(0.874)
Grade 4	307	5.867*	-0.040	-4.983*	-0.861
		(2.684)	(0.851)	(2.404)	(0.792)
Grade 5	308	4.796*	-0.729	-2.949	-1.081
		(2.349)	(1.190)	(1.634)	(0.602)
Grade 6	300	4.288	-0.598	-2.508	-1.157
		(2.968)	(0.902)	(1.743)	(1.040)
Grade 7	301	3.423	0.958	-2.940	-1.549
		(2.908)	(1.104)	(1.902)	(1.386)
Grade 8	300	3.422	0.734	-2.531	-1.671
		(3.309)	(0.975)	(1.863)	(1.303)
Grade 10	360	6.960	-3.123	-2.693	-1.139
		(3.858)	(1.969)	(1.451)	(0.799)
White Students					
All Grades	282	7.846**	-1.821	-4.384**	-1.593*
		(2.895)	(1.161)	(1.563)	(0.635)
Grade 3	308	3.478	-0.748	-1.510	-1.120*
		(2.556)	(0.988)	(1.421)	(0.486)
Grade 4	307	5.246*	0.020	-5.829**	-0.827
		(2.333)	(0.600)	(2.157)	(0.489)
Grade 5	308	7.277**	-1.529	-4.527**	-1.296
		(2.539)	(0.767)	(1.657)	(0.739)
Grade 6	300	5.146	-0.909	-3.021	-1.182
		(2.610)	(0.811)	(1.606)	(0.636)
Grade 7	301	4.472	-2.014	-2.357*	-0.156
		(2.497)	(1.851)	(0.963)	(0.290)
Grade 8	300	4.086	0.682	-3.002	-1.701
		(2.997)	(0.855)	(1.803)	(1.042)
Grade 10	360	6.675	-3.648	-2.459	-0.593
		(3.422)	(2.724)	(0.680)	(0.287)
Control District Average	809	38.241	34.974	19.934	6.878
		(17.516)	(8.650)	(9.616)	(4.864)

Standard errors in parentheses
*p<0.05, ** p<0.01, *** p < 0.001

NOTES: There is a unit of observation for each district-grade-year combination. "All grades" is counted as its' own grade category rather than as a sum of the other grades. The academic achievement analysis period is from the 2007-'08 school year to the 2013-'14 school year for grades 3 through 8. The academic achievement period for grade 10 is from 2007-'08 through 2017-'18 because in the NextGen test wasn't implemented until the 2018-'19 for grade 10. Control district average refers to the average achievement of all students in the matched sample used for grades 3-8 analysis in all years of the sample period. The sample period excludes the 2014-'15 and 2015-'16 school years due to the implementation of an experimental exam in these years. The full specification includes school year fixed effects and robust standard errors clustered at the district-level.

TABLE 6 – Effect of Metco on NextGen ELA Achievement

	Number of Observations	% Exceeds Expectations	% Meets Expectations	% Partially Meeting Expectations	% Not Meeting Expectations
All Students					
Grade 3	132	2.408	2.871	-4.508	-0.867
		(2.519)	(2.043)	(3.145)	(0.635)
Grade 4	132	5.414*	0.617	-4.825	-1.302
		(2.282)	(1.849)	(3.025)	(0.720)
Grade 5	132	5.119**	1.532	-5.711	-0.984
		(1.759)	(2.113)	(3.092)	(0.786)
Grade 6	129	2.390	-0.385	-1.254	-0.655
		(2.517)	(1.434)	(2.527)	(0.713)
Grade 7	129	1.626	0.557	-0.968	-1.050
		(1.841)	(1.748)	(2.336)	(0.954)
Grade 8	129	2.599	0.666	-2.235	-1.211
		(2.891)	(1.668)	(2.513)	(0.890)
Grade 3-8 (full sample)	135	3.445	1.174	-3.468	-1.239
		(1.812)	(1.451)	(1.239)	(0.741)
Grade 3-8 (subsample)	121	4.430*	2.356	-4.946	-1.915
		(1.905)	(1.807)	(2.713)	(0.950)
White Students					
Grade 3	132	2.611	2.821	-4.933	-0.732
		(2.387)	(1.849)	(2.726)	(0.471)
Grade 4	132	5.433*	1.241	-5.746*	-0.862
		(2.066)	(1.581)	(2.625)	(0.536)
Grade 5	132	4.727**	1.494	-5.660	-0.448
		(1.609)	(1.828)	(2.935)	(0.433)
Grade 6	129	2.523	-0.394	-1.672	-0.516
		(2.461)	(1.326)	(2.489)	(0.569)
Grade 7	129	1.553	1.394	-1.579	-1.239*
		(1.638)	(1.725)	(2.300)	(0.596)
Grade 8	129	2.828	1.677	-3.397	-1.133
		(2.601)	(1.604)	(2.371)	(0.683)
Grade 3-8 (full sample)	135	3.445	1.174	-3.468	-1.239
		(1.812)	(1.451)	(2.386)	(0.741)
Grade 3-8 (subsample)	121	4.192*	2.185	-4.829*	-1.420*
		(1.588)	(1.363)	(2.281)	(0.558)
Black Students					
Grade 3-8 (subsample)	121	0.120	-3.060	3.086	-0.271
		(1.360)	(3.504)	(3.636)	(2.077)
Control District Average	3,918	10.906	46.007	34.964	8.156
		(9.974)	(14.892)	(15.691)	(8.886)

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

NOTES: There is a unit of observation for each district-grade-year combination. The academic achievement period of analysis is the 2016-'17 school year through the 2018-'19 school year. The subsample was constructed by removing district-year observations that are not reported for the Black student subgroup from the full sample to ensure that the same district-year observations were analyzed for each group of students. The full specification includes school year fixed effects and robust standard errors clustered at the district-level. The control district average refers to the average achievement of all students in grades 3-8 in control districts.

TABLE 7 – Effect of Metco on NextGen Math Achievement

	Number of Observations	% Exceeds Expectations	% Meets Expectations	% Partially Meeting Expectations	% Not Meeting Expectations
All Students					
Grade 3	132	1.894	3.102*	-4.419	-0.695
		(2.710)	(1.534)	(3.081)	(0.782)
Grade 4	132	3.160	5.310*	-7.006*	-1.654
		(2.285)	(2.252)	(3.392)	(1.039)
Grade 5	132	3.505	2.844	-5.279	-1.127
		(2.069)	(2.710)	(3.445)	(0.996)
Grade 6	129	1.286	0.961	-0.920	-1.292
		(2.691)	(1.856)	(3.137)	(0.872)
Grade 7	129	0.680	2.093	-1.349	-1.387
		(3.017)	(1.529)	(3.071)	(0.923)
Grade 8	129	-0.440	1.837	-0.559	-0.861
		(3.281)	(1.905)	(3.123)	(0.987)
Grade 3-8 (full sample)	135	1.882	2.902	-3.644	-1.251
		(2.397)	(1.697)	(2.995)	(0.876)
Grade 3-8 (subsample)	121	3.355	3.089	-4.938	-1.672
		(2.502)	(1.824)	(2.998)	(1.081)
White Students					
Grade 3	132	1.944	3.678*	-4.621	-0.930
		(2.292)	(1.531)	(2.758)	(0.588)
Grade 4	132	2.948	6.333*	-7.976*	-1.444
		(1.858)	(2.371)	(3.163)	(0.858)
Grade 5	132	3.271*	3.863	-6.246*	-1.014
		(1.462)	(2.537)	(3.068)	(0.755)
Grade 6	129	1.321	1.387	-1.827	-0.894
		(1.960)	(1.742)	(2.692)	(0.584)
Grade 7	129	1.118	2.701	-2.397	-1.320*
		(2.337)	(1.458)	(2.797)	(0.619)
Grade 8	129	0.0731	1.892	-1.651	-0.456
		(2.538)	(1.868)	(2.944)	(0.662)
Grade 3-8 (full sample)	135	1.881	3.576*	-4.339	-1.101
		(1.711)	(1.634)	(2.560)	(0.572)
Grade 3-8 (subsample)	121	4.192*	2.185	-4.829*	-1.420*
		(1.240)	(3.749)	(3.075)	(2.213)
Black Students					
Grade 3-8 (subsample)	121	-0.892	-3.268	5.130	-0.937
		(1.240)	(3.749)	(3.075)	(2.213)
Control District Average	3,913	12.859	43.638	34.816	5.262
		(12.419)	(14.869)	(16.582)	(3.186)

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

NOTES: There is a unit of observation for each district-grade-year combination. The academic achievement period of analysis is the 2016-'17 through 2018-'19 school years. The subsample was constructed by removing district-year observations that are not reported for the Black student subgroup from the full sample to ensure that the same district-year observations were analyzed for each group of students. The full specification includes school year fixed effects and robust standard errors clustered at the district-level. The control district average refers to the average achievement of students in grades 3-8 in the control districts.

TABLE 8 – Robustness Check Impact on District-Level Estimates

	N	%Advanced	%Proficient	%Needs Improvement	%Warning/Failing
ELA					
Check 1)	282	3.390	-1.396	-1.270	-0.699**
		(1.962)	(1.130)	(0.810)	(0.231)
Check 1a)	282	3.390	-1.396	-1.270	-0.699*
		(2.153)	(1.251)	(0.897)	(0.256)
Check 2)	228	4.915	-1.922	-1.981	-0.944
		(2.969)	(1.701)	(1.404)	(0.534)
Check 3)	282	2.459	-1.371	-0.708	-0.387
		(2.220)	(1.372)	(0.949)	(0.236)
Math					
Check 1)	282	4.219	-0.596	-2.527*	-1.026*
		(2.184)	(0.972)	(1.033)	(0.502)
Check 1a)	282	4.219	-0.636	-2.545*	-1.020
		(2.418)	(1.076)	(1.143)	(0.556)
Check 2)	228	5.903	-1.444	-2.937	-1.401
		(3.516)	(1.418)	(1.836)	(1.022)
Check 3)	282	3.306	-0.804	-1.892	-0.520
		(2.482)	(0.871)	(1.487)	(0.647)

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

NOTES: There is a unit of observation for every district-year combination. This table shows estimates of Metco's district-level effect on the Legacy with the three robustness checks incorporated. Checks 1, 1a, 2, and 3 correspond to the baseline covariate controls, covariate and school year interaction check, imposition of common support check, and the pre-period academic achievement check respectively. I performed the robustness checks for all students and White students at each grade level for both the Legacy and NextGen tests, but only show the district-level results as I found a similar pattern across all. As with the main results, this specification includes school year fixed effects and robust standard errors clustered at the district-level.

12 References

American Psychological Association. n.d. "Education and Socioeconomic Status Factsheet." Accessed April 17, 2022. https://www.apa.org/pi/ses/resources/publications/education#:~:text=SES%20and%20Academic%20Achievement,compared%20with%20higher%20SES%/20communities.

Angrist, Joshua, and Kevin Lang. 2004. "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program." *American Economic Review* 94 (5): 1613-34. doi:10.1257/0002828043052169.

Austin, Peter C. 2011. "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." *Multivariate Behavioral Research* 46 (3): 399–424. https://doi.org/10.1080/00273171.2011.568786.

Batson, Ruth M. & Robert C. Hayden. 1987. *Special Preview Copy of A History of METCO.* Archives and Special Collections, Northeastern University Library, Boston, MA. http://hdl.handle.net/2047/D20287308.

Billings, Stephen B., David J. Deming, and Jonah Rockoff. 2014. "School Segregation, Educational Attainment, and Crime: Evidence from the End of Busing in Charlotte-Mecklenburg." *The Quarterly Journal of Economics* 129 (1): 435-76. doi:10.1093/qje/qjt026.

Card, David, and Jesse Rothstein. 2007. "Racial segregation and the black-white test score gap." *Journal of Public Economics* 91 (11-12): 2158-84. doi: 10.1016/j.jpubeco.2007.03.006.

Delmont, Matthew. 2016. "The Lasting Legacy of the Busing Crisis." *The Atlantic*, March 29, 2016. https://www.theatlantic.com/politics/archive/2016/03/the-boston-busing-crisis-was-never-intended-to-work/474264/.

"Desegregation Busing." n.d. *Boston Research Center*. Accessed April 17, 2022, https://bostonresearchcenter.org/projects_files/eob/single-entry-busing.html.

Garrido, Melissa M., Amy S. Kelley, Julia Paris, Katherine Roza, Diane E. Meier, R. Sean Morrison, and Melissa D. Aldridge. 2014. "Methods for Constructing and Assessing Propensity Scores." *Health Services Research* 49 (5): 1701–20. https://doi.org/10.1111/1475-6773.12182.

Guryan, Jonathan. 2004. "Desegregation and Black Dropout Rates." *American Economic Review* 94 (4): 919-43. doi:10.1257/0002828042002679.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2009. "New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement." *Journal of Labor Economics* 27 (3): 349-83. https://doi.org/10.1086/600386.

Hanushek, Eric A., and Lori L. Taylor. 1990. “Alternative Assessments of the Performance of Schools: Measurement of State Variations in Achievement.” *The Journal of Human Resources.* 25 (2): 179-201. doi: 10.2307/145753.

Hoxby, Caroline. 2000. “Peer Effects in the Classroom: Learning From Gender and Race Variation.” National Bureau of Economic Research Working Paper 7867. http://www.nber.org/papers/w7867.

Leuven, E. and B. Sianesi. (2003). "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing". http://ideas.repec.org/c/boc/bocode/s432001.html, version 4.0.12.

Lundstrom, Samuel. 2017. “The Impact of Family Income on Child Achievement: Evidence From the Earned Income Tax Credit: Comment.” *American Economic Review* 107 (2): 623–28. https://doi.org/10.1257/aer.20150988.

Mantil, Ann. 2018. “Evaluating the Impact of Nontraditional School Choices.” Doctoral diss., Harvard Graduate School of Education. https://dash.harvard.edu/handle/1/37679891.

Mar, Don. 2018. “Racial Diversity and Academic Test Scores in Public Elementary Schools.” *Applied Economics Letters* 25 (11): 768-71. https://doi-org.ezproxy.amherst.edu/http://www.tandfonline.com/loi/rael20.

Massachusetts Department of Elementary and Secondary Education. 2005-2014. “Achievement and Growth by District.” Aggregate File Database. Retrieved from https://www.doe.mass.edu/infoservices/research/default.aspx?id=F9C9F791B2FE4596A3DC2A4338086114E7454C3E.

Massachusetts Department of Elementary and Secondary Education. 2016-2019. “MCAS Achievement and Growth (NextGen) by District.” Aggregate File Database. https://www.doe.mass.edu/infoservices/research/default.aspx?id=F9C9F791B2FE4596A3DC2A4338086114E7454C3E.

Massachusetts Department of Elementary and Secondary Education. 2006. “2006-07 Teacher Data - All Teachers.” Retrieved from https://profiles.doe.mass.edu/statereport/teacherdata.aspx.

Massachusetts Department of Elementary and Secondary Education. 2003-2021. “Metco Enrollment by District 2003-2021”. [Dataset].

Massachusetts Department of Elementary and Secondary Education. “Understanding the Next-Gen MCAS Powerpoint Presentation.” PowerPoint presentation, Malden, MA October, 2017. https://www.doe.mass.edu/mcas/parents/understand-nextgen.html.

Massachusetts Department of Elementary and Secondary Education. 2018. *Metco Legislative Report 2017.* Malden, MA. https://www.doe.mass.edu/infoservices/research/default.aspx?id=F9C9F791B2FE4596A3DC2A4338086114E7454C3E.

Massachusetts Department of Elementary and Secondary Education. 2020. "Researcher's guide to Massachusetts state education data." Malden, MA. 2020. https://www.doe.mass.edu/research/researchers-guide.docx.

"METCO." n.d. *Springfield Public Schools*. Accessed April 19, 2022. https://www.springfieldpublicschools.com/departments/pace/resources/m_e_t_c_o.

Reardon, Sean F. 2016. "School Segregation and Racial Academic Achievement Gaps." *RSF:The Russell Sage Foundation Journal of the Social Science* 2 (5): 34-57. doi: 10.7758/RSF.2016.2.5.03.

Rivkin, Steven G. 2000. "School Desegregation, Academic Attainment, and Earnings." *The Journal of Human Resources* 35 (2): 333-46. doi: 10.2307/146328.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score In Observational Studies for Causal Effects." *Biometrika* 70 (1), 41-55. doi: https://doi.org/10.1093/biomet/70.1.41.

Sacerdote, Bruce. 2011. "Peer Effects in Education: How Might They Work, How Big Are They And How Much Do We Know Thus Far?" In *Handbook of the Economics of Education*, 3:249–77. Elsevier. doi: https://doi.org/10.1016/B978-0-444-53429-3.00004-1.

Semuels, Alana. 2019. "The Utter Inadequacy of America's Efforts to Desegregate Schools." *The Atlantic,* April 12, 2019. https://www.theatlantic.com/education/archive/2019/04/boston-metco-program-school-desegregation/584224/.

Sharfenberg, David. 2020. "Massachusetts' public schools are highly segregated. It's time we treated that like the crisis it is." *The Boston Globe*, December 11, 2020. https://www.bostonglobe.com/2020/12/11/opinion/massachusetts-public-schools-are-highly-segregated-its-time-we-treated-that-like-crisis-it-is/.

Smith, Jefferey. 2009. "Economics 675 – Matching Lectures." University of Michigan, Ann Arbor, Michigan. December 2009.

Tam, Mo Yin S., and Gilbert W. Bassett. 2004. "Does Diversity Matter? Measuring the Impact of High School Diversity on Freshman GPA." *Policy Studies Journal* 32 (1): 129-43. doi: https://doi.org/10.1111/j.0190-292X.2004.00056.x.

The Metropolitan Council For Educational Opportunity. n.d. "About METCO." Accessed April 18, 2022. https://metcoinc.org.

Turner, Julia. 2019. "The Impact of Resegregation on the Racial Achievement Gap: Insights from Seattle." Bachelor's thesis., Amherst College.

U.S. Census Bureau's Population Division. 2000. "The Census (2000) School District Tabulation (STP2)". Institute of Education Sciences, National Center for Education Statistics. Retrieved from https://nces.ed.gov/programs/edge/TableViewer/census/2000.

U.S. Census Bureau. 2006. "School District Estimates for 2006." Small Area Income and Poverty Estimates (SAIPE) Datasets. Retrieved from https://www.census.gov/data/datasets/2006/demo/saipe/2006-school-districts.html.

U.S. Department of Education. 2006. "Local Education Agency (School District) Universe Survey Data, v.1c." Common Core of Data. Institute of Education Sciences, National Center for Education Statistics. Retrieved from https://nces.ed.gov/ccd/files.asp#Fiscal:2,LevelId:5,SchoolYearId:21,Page:1.

Vaznis, James. "Metco will move to a lottery system to choose students." *The Boston Globe*, May 6, 2019. https://www.bostonglobe.com/metro/2019/05/06/state-approves-new-rules-for-metco-admissions/qNh0314pMyZ4j321MMAxDM/story.html.