The Impact of Increased Exposure of Diversity on Suburban Students' Outcomes: An Analysis of the METCO Voluntary Desegregation Program

Elizabeth Setren*

October 4, 2022

DO NOT CITE OR CIRCULATE WITHOUT PERMISSION OF AUTHOR.

Abstract

Over sixty years following Brown vs. Board of Education, racial and socioeconomic segregation and lack of equal access to educational opportunities persist. Across the country, voluntary desegregation busing programs aim to ameliorate these imbalances and disparities. A longstanding Massachusetts program, METCO, busses non-white K-12 students from Boston and Springfield, Massachusetts to 38 suburban districts that voluntarily enroll urban students. Supporters of the program cite that it prepares students to be active citizens in our multicultural society. Opponents question the value of the program and worry it may have a negative impact on suburban student outcomes. I estimate the causal effect of exposure to diversity through the METCO program by using three types of variation: difference-in-difference analysis of schools stopping and starting their METCO enrollment, sporadic changes in which cohorts have METCO peers, and instrumental variable analysis of space availability for METCO students. I analyze the impact of exposure to increased student diversity on suburban students' outcomes, including academic performance, classroom behavior, and attendance. All methods rule out substantial test score, attendance, or suspension effects of having METCO peers. Classroom ability distribution remains similar when METCO programs start and stop. Ongoing work analyzes the impact on civic, social, and exposure to diversity outcomes, as well as the effect on urban participants.

*Tufts University, Department of Economics, Medford, MA 02155, elizabeth.setren@tufts.edu. Thanks to seminar participants at CESifo Education Program, AEFP 2021, Syracuse University, Vanderbilt University, Harvard Program on Education Policy and Governance, Harvard Kennedy School, and Tufts University for helpful feedback. I am grateful to The Boston Foundation, the Institute for Education Sciences, The Spencer Foundation, The Russell Sage Foundation, and Tufts University Faculty Research Fund for providing generous financial support for this project. Many thanks to Savannah Kochinke and Elizabeth Pancotti for providing superb research assistance. This research would not have been possible without a major digitization effort: thanks to Rebecca Rocha, Jaeho Kim, Evan Newell, Julia Bradley, David Zhao, Madeira Thayer, Vasudha Ramakirshna, Rebecca Gertler, Elena Chavez, Ashley Lamoretti, Karen Trop, Maya Hamberg, Hannah Mirviss, Ace Haws, Denise Wilson, and Shirley Porcena for digitizing the historical records. Thanks also to Eryn Heying for administrative support. Special thanks go to Matthew Deninger, Carrie Conaway, Cliff Chuang, and the staff of the Massachusetts Department of Elementary and Secondary Education for data, suggestions, and assistance. I am incredibly grateful to Milly Arbaje-Thomas, Charles Walker, Jean McGuire, John Shandorf, Wilmary Tejeda, Kim Houston, Deborah Ward, Colin Stokes, Kristen Fumarola, Jeffrey Welch, and the rest of the METCO staff for sharing their expertise, feedback, historical records, and data.

Over 60 years following Brown v. Board of Education, racial and socioeconomic segregation and lack of equal access to educational opportunities persist. The problem is growing: the proportion of K-12 public schools with over 75 percent of poor students of color has nearly doubled since 2001 (Government Accountability Office, 2016). Research links school segregation to widening disparities in academic achievement (Card and Rothstein, 2007; Vigdor and Ludwig, 2007).

Across the country, voluntary desegregation busing programs aim to ameliorate these imbalances and disparities by sending students of color from urban school districts to predominantly white, suburban schools.[1] A longstanding Massachusetts program, the Metropolitan Council for Educational Opportunity (METCO) serves as a national model for other desegregation programs throughout the country (Eaton, 2001). Founded in 1966, METCO buses students of color from Boston and Springfield, Massachusetts to 38 suburban K-12 school districts that voluntarily elect to enroll urban students. Currently, over 3,150 Boston and 150 Springfield students attend suburban districts through the METCO program. The state-funded program aims to promote diversity and cultural competency by reducing racial isolation in suburban and urban districts and to increase access to high performing schools for urban students.

Supporters of METCO, and other efforts to increase school diversity, proport that it has positive peer effects through exposing students to different cultures, backgrounds and views. Research suggests reduced racial isolation may help students overcome racial stereotypes, learn from other cultures, and reduce prejudice (Godsil et al., 2014; Zebrowitz, White and Wieneke, 2008). Others express concern that there may be negative peer effects on academics and behavior from increased racial and socioeconomic diversity (McGuirk, 2019; Town of Brookline Board of Selectmen, 2014, Muriel Cohen, 1990, Doreen Iudica Vigue, 1999). More generally, parents' perception of school quality and reputation is highly influenced by the share of students of color (Ellen, 2000 and Wells, 2015). These concerns, whether implicit or explicit, counteract integration efforts and worsen school segregation.

This paper provides causal estimates on whether increased exposure of urban students of color has academic or behavioral peer effects on suburban students using three identification strategies. Each strategy uses state administrative education data from 2001 - 2020 and future work will incorporate historical data back to 1966. The first uses fluctuations in METCO enrollment across grade cohorts within schools to estimate the causal impact of having METCO students as peers in a difference-in-differences analysis. The second focuses on schools that switch whether they enroll new METCO policy frequently, and compares cohorts with and without METCO peers within school. The final method harnesses the fact that classroom space constraints play a role in determining the number of METCO students a school accepts each year in a method similar to Angrist and Lang (2004)'s adaptation of Maimonides' rule.

The results teach us the effects of increased integration both a large city, Boston, and a mid-sized city, Springfield and across nearly two decades (and follow-up work will include over fifty years). This paper builds upon Angrist and Lang (2004)'s study of the Brookline school district's METCO program during 1994 - 2000 and the broader sample enables heterogeneity and mechanisms investigations. This work also contributes to the literature on the effects of desegregation and more broadly, open enrollment and school choice policies (Johnson, 2015; Reber, 2010; Bergman, 2018, Bruhn, 2021).

All methods rule out substantial test score, attendance, or suspension effects of having any METCO peers in the cohort or an additional METCO student per class. I also find that having METCO peers widens the classroom 90th - 10th percentile ability distribution by four percent in Math and does not change the distribution in English.

The conclusive null effects of urban students on the suburban students' outcomes suggests that the concerns of negative academic or behavioral effects from increases urban students of color in Boston and Springfield-area suburban schools are unwarranted. Continued work will estimate the impact of METCO on participants using waitlist instruments and incorporate historical data and extend the results to longer-term outcomes, including college enrollment and graduation, employment and earnings, civic engagement, criminal justice outcomes, marriage and fertility and housing outcomes.

[1] East Palo Alto, Omaha, St. Louis, Minneapolis, Milwaukee, and Rochester each have voluntary desegregation busing programs.

1 Background

Origin of the METCO Program

METCO formed in the wake of intense opposition to integrating Boston Public Schools and improving school quality for Boston's black students. Inadequate school facilities and resources, overcrowding, and defacto segregation led to a series of protests and school boycotts from black parents, students, and community members. A State Board of Education commission to study school segregation in Boston concluded that the racial imbalance in Boston harmed both black and white students, but the recommendations of the commission were denounced by the Boston School Committee, who refused to take action.

Faced with opposition from the School Committee, Black parents forged their own integration program called Operation Exodus – bussing over 400 students from overcrowded schools to under-enrolled schools outside of predominantly black neighborhoods. This evolved into the METCO program. The school committees of Boston suburbs Brookline, Lexington, Newton, and Wellesley voted to accept black students from Boston in January 1966. Four other districts followed suit and in the Fall of 1966, 220 K-11 students from the city of Boston enrolled in suburban schools. Initially intended as a temporary program until desegregation was implemented in Boston Public Schools, METCO expanded as new districts elected to participate. Meanwhile, court-ordered integration in Boston sparked intense opposition and a series of within-city desegregation bussing plans.

To join the METCO program, the town's school committee, town council, or equivalent voted to participate. Then METCO decided whether to accept the district. Several districts were turned down because there was widespread opposition to the METCO program and the town approved METCO by only a small margin.[2] METCO accepted new districts until 1975. Except for four districts that ended their METCO programs, after joining METCO, districts have continuously enrolled METCO students.[3] The city of Springfield started its own METCO program with four nearby suburbs. It currently enrolls 150 K-12 students a year. Figure A.1 shows the growth in METCO enrollment as new districts joined and relatively stable enrollment of over 3,000 students since the 1980s.

Once a student is accepted into a METCO suburban district, they can remain enrolled in that district until they graduate. Suburban METCO districts vary in the degree of services they provide urban students. Each district has a METCO Director who oversees the program and enrollment process. Some districts have additional services to support students academically, socially, and emotionally and to increase the inclusion of METCO students in the school district. such as late buses, so that students can participate in after school activities, tutoring, access to social workers and counselors, and social programs to foster relationships with students and families at the school. In addition, METCO urged the suburban districts to have inclusive curriculums that are free from racial and gender stereotypes and include African and African-American history and culture in all disciplines.[4] METCO also encouraged the districts to hire black teachers and staff.[5] Today, METCO is working with suburban districts to help them strive to be antiracist institutions.

Reasons for Suburban Support and Opposition

Districts joined METCO to increase the diversity of their schools and prepare their students to interact with people from different backgrounds. Both METCO and districts cited this reason, saying it "would broaden the experience"[6] of students and create "a new learning experience for suburban children."[7] Moral responsibility or charity also motivated some suburbanites to participate. METCO pushed back against this reason because it ignored the benefit urban students brought to the suburban schools and stereotyped METCO students as all poor[8]. While the diversity in the suburban schools has increased since they joined the METCO program in the 1960s and 1970s, they still remain predominantly white and the motivation to

[2] METCO rejected Beverly, Georgetown, Randolph, and Winchester in 1974 and Winchester was also rejected in 1967.

[3] Milton, Rockland, Hamilton-Wenham, and Framingham ended their METCO programs.

[4] "Guidelines for Metropolitan Boston Communities. . . ," 1970, METCO Archives.
"METCO Handbook," 1976, METCO Archives.

[5] "Guidelines for Metropolitan Boston Communities. . . ," 1970, METCO Archives.

[6] "Supports School Board," Lexington Minute-Man, January 27, 1966.

[7] Bureau of Equal Educational Opportunity, "METCO Handbook," Massachusetts Department of Education, 1976, Box 1, File 39, METCO Archives.

[8] Ann Tutino, "Young Bars Special Role for Negroes: Treat Them As Any Others, School Chief Declares."

increase school diversity remains today. There is policy discussion around expanding the seat offerings in METCO and eight new districts have inquired about joining the program.[9]

Opponents to the METCO program focus on cost concerns and potential negative spillover effects on suburban students. The state of Massachusetts funds the METCO program (more details below). Even though local property taxes did not fund METCO, residents expressed concern that property taxes would increase to pay for the program and that they didn't consent to have their state tax dollars spent on the program.[10] Particularly during the economic downturn in the 1970s, some suburbanites opposed spending their tax dollars for limited integration. When Newton faced declining enrollments in the 1970s, a contingent advocated for closing a school and consolidating enrollment to save money instead of accepting METCO students into empty seats.[11]

Connected to the cost concerns, opponents argued that METCO would increase class sizes and reduce the amount of individualized attention their students would receive from teachers. They also worried that urban students would increase the behavioral disruptions in the classroom and lower the average test scores of the school. In turn, this change in school quality could have a negative impact on suburban students' outcomes, perceived school quality, and in turn, housing values.[12]

Others argued that the program was discriminatory because it focused on urban students of color instead of including all urban students.[13] Some contended that the program burdened students with the long bus rides, but ignored the fact that their families willingly signed up for the program knowing the length transit time.[14] Public discourse rarely discussed whether the housing discrimination and the resulting residential segregation should be addressed, which would in turn reduce school segregation without the need for bussing.[15]

Funding

Districts receive $8,900 on average for each METCO student, though this amount varies by district and year.[16] Funding comes from a state METCO grant of approximately $5,000 per student and from state education funding because METCO students count towards the districts' total enrollment which is used to calculate state education aid.[17] Since the state aid is not itemized into funding for METCO and resident students, it is not transparent to districts how much funding they receive from the METCO program beyond the grant.[18]

Since METCO students typically fill empty classroom seats instead of causing new classrooms to be formed, the districts pay the marginal cost of a student. Under this model, it's possible that districts do not incur costs of hiring new classroom teachers. Districts cover the cost of transportation, special education services, instructional materials and costs, salaries for a METCO program director, and salaries for other

[9]Apfelbaum, Katherine and Ardon, Ken, "Expanding METCO and Closing Achievement Gaps." Pioneer Institute White Paper No. 129. March 2015.

Joseph, Stephanie. "After George Floyd, More Suburbs Express Interest in Joining METCO." Learning Curve, WGBH. October 6, 2021. https://www.wgbh.org/news/education/2021/10/06/after-george-floyd-more-suburbs-express-interest-in-joining-metco Accessed February 21, 2022.

[10]Chanoux, Laura, "From the City to the Suburbs: School Integration and Reactions to Boston's METCO Program. Senior Honors Thesis. March 30, 2011.

[11]A. Landsman, "Questions for C.E.E.N.," Newton Graphic, November 14, 1974, Box 44, File 16, METCO Archives.

Michael R. LeConti, "Open letter to School Committeeman ward 3 Gerald Byrne," Newton Villager, November 28, 1974, Box 44, File 15, METCO Archives.

[12]Setren (2010) finds no evidence of declines in housing prices when districts begin or expand their METCO enrollments.

[13]"House kills two bills on METCO program," Boston Globe, May 1, 1975, METCO Archives.

John W. Cole, "About Our Schools," Lexington Minute-Man, August 18, 1966, Box 53, File 30, METCO Archives.

[14]Chanoux, Laura, "From the City to the Suburbs: School Integration and Reactions to Boston's METCO Program. Senior Honors Thesis. March 30, 2011.

"Scatteration' won't help," Boston Herald, January 1966, Box 44, File 26, METCO Archives.

[15]Chanoux, Laura, "From the City to the Suburbs: School Integration and Reactions to Boston's METCO Program. Senior Honors Thesis. March 30, 2011

[16]Apfelbaum, Katherine and Ardon, Ken, "Expanding METCO and Closing Achievement Gaps." Pioneer Institute White Paper No. 129. March 2015.

[17]The state funding formula uses districts' prior year student enrollment and allocates more funding for certain grade levels, English Language Learners, and low-income students.

[18]Apfelbaum, Katherine and Ardon, Ken, "Expanding METCO and Closing Achievement Gaps." Pioneer Institute White Paper No. 129. March 2015.

program support staff, like counselors.

Literature

Qualitative work on METCO interviews and surveys parents of METCO students and adults that participated in METCO (Boardman and Brandt, 1968; Clarke, 1975; Orfield et al., 1997; Eaton, 2001; Elliott, 1998). A study comparing METCO students with a small number of non-participating siblings did not find benefits to academic achievement or race relations Armor (1972). While the literature provides rich insights into students' experiences in METCO, it is unable to identify the causal effect of METCO without proper control groups, complete follow-up outcomes, or sufficient controls Angrist and Lang (2004). Mantil (2022) looks at the relationship between whether a grade cohort has more potential space for METCO students (based on total METCO district enrollment), and the outcomes of seven birth cohorts of METCO applicants. Yearly variation in METCO enrollment may be too coarse a strategy to precisely identify treatment effects. Mantil also compares referred and non-referred METCO applicant outcomes for high school graduation and college enrollment for three high school cohorts. In a related study, I incorporate all of the applicant and referral data from 1966 to present and use two-stage least squares methods to estimate the causal impact of METCO on its participants for a broad range of K-12, college, and longer-run outcomes (Setren 2022).

Very little work has been done on the impact of METCO on the suburban districts. Several studies have found a negative correlation between the percentage of students of color in schools and classes and academic performance, especially for minority students Lazear (2001). However, since the percentage of students of color could serve as a proxy for economic and social differences, it is unclear how this mechanism might work. Angrist and Lang (2004) found that even though METCO students have significantly lower test scores, there is little evidence of METCO students impacting the performance of their non-METCO classmates. In contrast to this earlier work, this project will study a wider range of school districts and a richer set of outcomes, including the impact of classroom diversity on school culture and climate, academic and social inclusion, bullying, and teacher expectations.

This paper contributes to the effects of court-ordered and voluntary desegregation literature.[19] Similar to Angrist and Lang (2004), causal evidence points towards a null or limited effect of school integration on white students' outcomes. Johnson (2011) finds positive effects of court-ordered school desegregation on Black students' educational, career, earnings, and health outcomes with reduced likelihood of incarceration and no effect on white students. Bergman et al. (2018) which studies the impact of an inter-district desegregation program and finds positive effects on test scores and college enrollment, but increases in special education classification and arrests. Guryan (2004) finds no effect of court-ordered desegregation on white students' high school dropout rates. This paper causally estimates the impact of a high-profile voluntary desegregation program on suburban students. Using have rich data sources spanning over 50 years to investigate a wide-array of short and long-run outcomes and mechanisms. The project will also contribute to literature on school choice and school quality by increasing understanding of how school assignment and school diversity can impacts student outcomes.

2 Data

To study the impact of exposure to METCO peers on suburban students, this paper uses state administrative education records from 2001 - 2019. These data include detailed information for all students enrolled in Massachusetts public schools such as which schools they attend, whether they participate in METCO, demographics (including race, ethnicity, gender, free/reduced price lunch), special education status, English Language Learner status, standardized test scores, grade progression, attendance, suspensions, SAT and AP test taking and scores, qualifying for the Massachusetts state merit scholarship, passing the high school exit exam, and high school graduation, and students' self-reported aspirations for after high school. In addition, the data include postsecondary enrollment and degree completion for 2003 to present from the National Student Clearinghouse (NSC).

Unique student identifiers enable the linkage of students to their courses and teachers, providing data on which courses students take, their classroom peers, and their teachers' characteristics, including race/ethnicity,

[19] See Schofield (1991) and Schofield and Hausmann (2004) for detailed reviews of the literature.

years of experience, and certification. It also enables the calculation of average core class size.[20] The student-classroom-teacher linkage enables the estimation of teacher value-added. Personnel data also include counts of METCO-related staff, guidance counselors, and other specialists.

I use the following formula, adapted from Figlio et al. (2021), for METCO exposure to allow for flexible weighting of a student's proportion of peers from METCO across different grade levels:

$$a_{gjti} = \frac{\sum_{g' \leq g} Proportion.METCO_{ig'jt} e^{[1-(\lambda(g-g'))]}}{\sum_{g' \leq g} e^{[1-(\lambda(g-g'))]}} * 20 \quad (1)$$

When λ equals 0, the formula gives equal weight to the presence of METCO students in all grades — from grade 1 until outcome grade g. As the value of λ increases, more recent grades weigh more heavily into the exposure variable relative to the earlier grades. I multiply the weighted ratio by 20 so that the coefficient on a_{gjti} is the average number of METCO students grade g, school j, and cohort t if there were 20 students per class for ease of interpretation. Therefore, the regression coefficient on a_{gjti} is interpretted as the average effect of one additional METCO classmate for grades 1 through the outcome grade.

3 Descriptive Statistics

Resident students of METCO suburban schools are 86 percent white, score about a half standard deviation above the average Math and English test scores in the state, and only six percent qualify for free and reduced price lunch (see Table 1). Their urban peers from the METCO program have markedly different demographics: they are 79 percent black and 14 percent Latinx. While attending the suburban districts, METCO students score over 0.6 standard deviations lower than their suburban peers on standardized tests with an average of 0.12 standard deviations and 0.27 standard deviations below the state mean in Math and English respectively. Despite these lower test scores, METCO students score substantially higher than their peers in Boston Public Schools and Springfield Public Schools. Setren (2022) investigates how much of this difference is due to selection into the program by estimating the causal impact of METCO participation on urban students.[21]

Having peers with substantially lower test scores could negatively affect the suburban students if it changes the level of material taught in class, the amount of individual teacher attention they receive, and reduces the amount students learn from their peers. Classroom behavior is another mechanism for negative spillovers. Table 1 shows that METCO participants have almost three times higher suspension rates than their suburban peers and marginally lower attendance rates. If increased class absences or behavioral issues slow down instruction, then this could negatively affect suburban students. This paper investigates each of these channels to see whether there are negative spillovers on academic performance and behavioral outcomes. On average, more advantaged school districts opted into the METCO program. Table 1 shows that districts that did not join the METCO program have higher rates of economically disadvantaged students (23 percent versus 7 percent) and lower test scores (about 0.09 standard deviation above the state mean). Table 2 illustrates the positive selection using the following OLS regression:

$$y_{gjt,i} = \beta a_{gjt,i} + \zeta_g + \gamma_t + \epsilon_{gjt,i} \quad (2)$$

Similar to the descriptive statistics, Table 2 Column 2 shows that schools with a higher proportion of METCO students have higher test scores, higher attendance, and lower suspensions. However, controlling for district fixed effects, weakens the relationship between the proportion of METCO students and positive outcomes, though having METCO students is still strongly correlated with positive outcomes (Table 2, Column 3). Since controlling for individual districts lessens the relationship between METCO presence and suburban student outcomes, districts that participate in METCO were positively selected.

Once we restrict the sample to only suburban districts that opted into the program, we see that the districts that choose to accept relatively more METCO students also are positively selected on test scores:

[20] To determine the average class size in each school and grade, I get the class sizes for all core subject classes in the data. Then I define an individual student's class size as the largest class of any of their core classes. Finally, I calculate the average student class size for each school, grade, and year combination. Estimates are similar if instead average class size is the average size of all core classes in a school, grade, and year combination.

[21] For a detailed description of how urban students apply, get accepted, and enroll in the METCO program, see Setren (2022).

there is a muted relationship between proportion of METCO students in the school and suburban student test scores when we add district fixed effects (Table 2 Columns 5 and 6). The district fixed effects model shows a very small positive relationship between the proportion of METCO students and the suspension rate so either the districts that accept more METCO students are negatively selected on low suspension rates or having more METCO students could increase suspensions among resident students.

After controlling for individual districts, we see that having roughly one additional METCO student per classroom is linked with scoring 0.06 standard deviations higher on Math (much smaller than the 0.38 standard deviation difference without district controls). Controlling further for lagged school demographics, it appears that higher performing schools accept more METCO students since the estimates are further reduced, though still significant (Table 2 Column 7).[22] The positive relationship between METCO and suspension rates also appears for these models. Overall these regressions suggest the presence of substantial test score selection not just into which districts participate in METCO, but which districts and schools participate relatively more in the program.

After adding controls for individual student characteristics, there is still a positive relationship between the proportion of METCO students in a school and suburban students' test scores. Having one additional METCO student on average in a class is correlated with 0.026 and 0.036 standard deviations higher on Math and English, a 0.3 percent higher attendance rate and a 0.4 percent increase in the likelihood of suspension. These correlations are larger for suburban Black and Latinx students (see Table A.1).

The positive relationships between METCO and test scores and attendance could mean that the types of schools that enroll more METCO students also enroll suburban students that are pre-disposed to be high performing in unobservable ways. It could also reflect that the types of schools that enroll more METCO students also generate stronger test performance relative to other schools. It's also possible that METCO students generate a small positive academic effect and attendance effect on their suburban peers, particularly for students of color. Alternatively, these correlations could mask negative spillovers. The small positive relationship between suspensions and METCO students could suggest some negative spillovers.

4 Methodology

To remove the selection issues, this paper uses three quasi-experimental methodologies on three types of variations in METCO enrollment to identify the causal impact of METCO peers on suburban students' outcomes. The first method uses the start or end of METCO enrollment within schools in a difference in differences design. The second method compares cohorts within a school that have METCO students slightly older or younger cohorts without METCO peers, for schools that alternative between accepting and not accepting METCO students from cohort to cohort. The final method uses the fact that schools can accept more METCO students when there are more empty seats in classrooms. With an instrumental variable for predicted space in classrooms, I estimate the impact of having an additional METCO student in the classroom. The first two methods estimate the effect of having any METCO peers, while the third method estimates the average effect of an additional METCO peer.

Decisions of whether and how many METCO students to enroll are made by district and school leaders in the Spring. These decisions could be influenced by the projected space in classrooms, budget changes and concerns, efforts to have a fixed number of total METCO students in the district across all grades, and idiosyncratic preferences of school and district leadership. It is difficult to know which factors contributed to each schools' enrollment levels without in depth interviews with current and past district leadership . For the identification strategies to be valid, we need the fluctuations in METCO enrollment to be unrelated to anticipated outcomes. That means in the Spring prior to enrolling in Kindergarten or first grade, parents do not influence how many METCO students their child's future school enrolls. In addition, school and district leaders do not select the number of METCO students in a cohort based on the expected outcomes of specific cohorts of students. This seems unlikely because the school does not have baseline academic outcomes for the incoming students.

The main analyses in this paper focus on elementary school entry which is the most common entry point for METCO.

[22]The lagged school demographics results are similar to a model with school fixed effects.

4.1 Approach 1: Difference in Differences

4.1.1 Empirics

I use fluctuations in METCO enrollment across grades within individual schools to estimate the causal impact of having METCO students as peers in a difference-in-differences analysis. Most schools in METCO receiving districts enroll METCO students at some point in time.[23] However, within the METCO-accepting schools, only 68.3 percent of the cohorts have at least one METCO student (See Figure 1 for percentages by district).[24] The cohorts without METCO students reflect either that a school has sporadic years where they pause their METCO enrollments or a school has stopped or not yet started their METCO enrollment. The former is the subject of the next section of this paper.

To utilize variation due to schools starting and stopping their METCO enrollment, I compare suburban student outcomes before and after schools start accepting METCO students. By using this within school comparison, I control for school-specific attributes that do not change over the sample period.

Over 90.9 percent of METCO-receiving districts accept METCO students in either kindergarten or first grade. After METCO students start in a suburban district, they can remain enrolled in that district until they graduate high school (as long as they continue to live in the city of Boston). Three districts start accepting METCO students in middle or high school, but otherwise a relatively small number of students start in the program after the entry grades. As such, the analysis compares cohorts of students that have at least one METCO student in first through fifth grade to cohorts that have no METCO students in any elementary school grade. Analyses of later outcomes will adjust this definition to incorporate the middle and high school years. Cohorts with METCO students in a subset of the first through fifth grades, but not all, are excluded from the analysis.

I define treatment schools as those that have a minimum of one year without METCO enrollment in the first grade followed by a minimum of one year with first grade METCO enrollment (results are robust to setting a two year minimum requirement). I include up to four years before the school enrolled METCO first graders and up to four years after in the analysis. For each schoolXyear-range that meets these sample requirements, I match it with its own set of comparison schools to adjust for time-varying trends. The comparison schools consist of the suburban schools in METCO-receiving districts that did not have METCO students in all of the cohorts during the treatment school's analysis sample years (up to five years before and four years after the METCO enrollment policy change).

I append the individual student-level data from each treatment school and its corresponding comparison schools to create a stacked dataset of all treatment and comparison groups. Since schools can appear as both a treatment school and a comparison school to multiple treatment schools, I include indicator variables **a**nd cluster standard errors on each treatment-comparison set. See Cengiz et al. (2019) and Deshpande and Li (2019) for other examples of stacked cohort differences-in-differences analysis.

This stacked difference-in-differences methodology corrects for the weighting problems created by staggered treatment timing discussed in Goodman-Bacon (2021). This methodology has similar properties as Sun and Abraham (2020) and Callaway and Sant'Anna (2020)**,** the other solutions to the weighting issues of staggered treatment timing (Baker, Larcker and Wang, 2021).

The sample includes all of the suburban resident students in the treatment and comparison schools and excludes METCO participants.

The following estimating equation identifies the causal impact of having METCO students in a grade cohort on resident suburban student outcomes:

$$y_i = \beta_0 + \beta_1 TreatmentSchool_{ic} + \sum_t D_i^t + \beta_2 TreatedCohort_i + \delta_j + X_i + \lambda_{tc} + \epsilon_{ic} \quad (3)$$

where y_i is the outcome of interest for student i. $TreatmentSchool_{ic}$ equals 1 if the student attended the school that undergoes METCO enrollment change c by switching to accepting METCO students. The cohort

[23]The exceptions include the elementary and middle schools in Foxborough, Natick, and Westwood which start accepting METCO students in middle or high school. Braintree, Lincoln, and Hampden-Wilbraham have at least one elementary school each that never accepts METCO students, but they have other elementary school(s) within the district that do.

[24]I calculate METCO representation in cohorts by using first grade enrollment numbers.

dummies, D_i^t control for the relative year t that student i enrolls in the first grade.[25] The cohort that attends first grade in the first year their school enrolls METCO students have $t = 1$. The following year's cohort has $t = 2$ and so on. The cohort right before the school accepts METCO students have $t = 0$ and two years before have $t = -1$.

$TreatedCohort_{ic}$ equals 1 if the student attended the school after policy change c, and therefore had METCO peers in their grade cohort. I include school fixed effects δ_j and individual baseline covariates X_i, including gender, race, immigrant status.[26] I control for the year that cohort t in enrollment change group c enters first grade with the indicator variables λ_{tc}. Since schools (and their enrolled students) can serve in multiple comparison groups, I cluster standard errors at the individual student level and the policy change subsample.

I repeat an analogous exercise for schools that change their enrollment policy from accepting METCO students, to no longer accepting them. Similar to the prior analysis, I include up to five years before the school enrolled METCO students and up to four years after. Suburban schools in METCO-receiving districts that had METCO enrollment in all of the cohorts during the treatment school's analysis sample years serve as the comparison schools. The estimating equation is identical except that policy change refers to ending new METCO enrollment.

The intensity of the treatment varies across the treated cohorts. On average, 4.1 percent of students come from the METCO program in the treated cohorts. The representation of METCO in treated cohorts ranges from 0.98 percent to 19.05 percent. This amounts to between 0.2 and 3.8 students in a 20 person class. Districts vary in the number of METCO students they accept, as shown in Figure 2. In the district with the smallest METCO program, 1.36 percent of the students in their treated cohorts come from the METCO program, compared to 8.86 percent in the largest METCO district.

4.1.2 Results

Table A.2 shows that both difference-in-differences analysis samples have broad coverage of districts, schools, and cohorts and similar baseline characteristics to suburban students overall (see Columns 2 and 3). In addition, over 73 percent of school districts serve in at least one treatment group, meaning they enrolled in first grade in the school within a four year range of when a school started or stopped new METCO enrollment.

Students in the difference-in-differences analysis samples have similar baseline characteristics as the full sample. There is no measure of baseline academic ability, but the pre-period third grade test scores of the difference-in-differences analysis samples are comparable to the full sample. Math and English test scores are about half a standard deviation above the state mean for the full sample and the analysis sample where ending the METCO program is the treatment. The analysis sample where starting METCO enrollment is the treatment has slightly lower pre-period test scores of 0.4 standard deviations above the state mean.

Table A.3 shows that treatment and comparison cohorts have similar baseline characteristics. Columns 2 and 5 show the coefficients from regressions of observed characteristics on being in a cohort following a METCO enrollment policy change. The models control for school fixed effects and the sample is the set of schools that experience policy changes. These results show that the cohorts before and after METCO enrollment policy changes have minimal demographic differences. While there are statistically significant differences for Black and immigrant, all differences are less than 1 percentage point, so the pre and post METCO program cohorts within schools have similar baseline characteristics. In addition, there was no differential pre-trend in baseline characteristics (see Figure A.2 and A.3).

Next, I compare the traits of treated and comparison schools. Columns 3 and 6 of Table A.3 show the regression of baseline characteristics on being in a treated school, controlling for the policy change subsample indicators and year. The results show that treated and comparison schools have similar demographic characteristics, with differences all less than 2 percentage points.

In addition to checking that the treatment and comparison groups have similar baseline characteristics, I check for common pre-trends in the outcomes of interest between the treated and comparison groups. I run

[25] If students repeat the first grade, I use the first attempt. If students switch schools during first grade, I use the school in a METCO-receiving suburb. If there are multiple schools in METCO-receiving suburbs, then I assign the student to the school where they attended the most days in first grade.

[26] I exclude controls that change over time such as special education status, free and reduced price lunch status, and English Language Learner status.

the following event study version of Equation 3 estimation:

$$y_i = \pi_1 TreatmentSchool_{ic} * \sum_t D_i^t + \sum_t D_i^t + \delta_j + \lambda_{tc} + X_i + \epsilon_{ic} \quad (4)$$

Figures 3 and 4 graph the results of this event study and shows comparable pre-trends across treatment and comparison groups for grade 3 through 5 test scores, suspension rates, and attendance rates. There are some small statistically significant differences in pre-trends for fifth grade suspension rate and fifth grade Math scores that I will address in robustness checks below. These small differences between treated and comparison groups are no longer significant when we consider correct standard errors for multiple testing.

Table 3 shows the results from the difference in differences regressions. Being in a cohort with METCO students has no significant effect on Math or English test scores. We can rule out Math effect sizes larger than plus or minus 0.03 standard deviations. For English, we can rule out effect sizes larger than plus or minus 0.035 standard deviations. This holds true when the treatment is a school starting METCO enrollment or ending METCO enrollment. There are also no significant effects of having METCO peers on suburban students' attendance or suspension rates: we can rule out effect sizes larger than plus or minus 0.2 percentage points. The event study graphs in Figure 3 show null effects of having METCO peers in each of the four cohorts who enter first grade after their school starts accepting METCO students. Similarly, Figure 4 shows no change in student outcomes in the four cohorts that enter after their school stops accepting METCO students.

While the event study graphs showed comparable pre-trends for most grade levels and outcomes, there were some small differences in pre-trends of the treatment and comparison groups for two fifth grade outcomes. Results are robust to narrowed comparison groups: requiring comparison schools are within the same district as the treated school and using synthetic control weights for the comparison schools. We fit the baseline characteristics of the treated cohorts to the year prior to the enrollment policy change (results are robust to assigning the synthetic control weights based on the two years prior to the policy change). Each of these methods yields similar results, though less precise point estimates due to the smaller sample size (see Table A.4).

4.2 Approach 2: Sporadic Differences

4.2.1 Empirics

Next, I use the fact that some schools have sporadic years with or without METCO cohorts. In the difference in differences analyses, I focused on schools with a stable METCO enrollment policy of accepting METCO students across several years that changed their policy. This approach focuses on schools that switch METCO enrollment policy year to year. Specifically, I define a treatment cohort as one that has METCO students and that school does not accept METCO in the cohort one year above and below. I run separate, analogous regressions where treatment is defined as not having METCO students and the cohort above and below in that school does have METCO students.

If the decision to have METCO in a cohort or not is orthogonal to the potential outcomes of those students, then the school's cohorts one year above and below serve as that treatment cohort's comparison group. For the METCO enrollment to be related to potential outcomes, district leaders would need to decide whether a cohort gets METCO or not based on the potential future test scores, behavior, and other outcomes of the incoming first graders. The district leaders decide on incoming METCO enrollment the Spring of the prior year, when they do not fully know which students will enroll. They do not have any prior information about the academic preparation or behavior of the incoming class to base their decision off of. We might also be concerned about parents changing their child's school based on METCO enrollment, but whether and how many METCO students will enroll in a given year is not transparent to parents. Since the decision is made in the Spring of the prior year and timing varies by school district, parents of incoming first graders are unlikely to have influence or know exactly when the decision is made.

The estimating equation is as follows:

$$y_i = \beta_0 + \beta_1 TreatedCohort_i + X_i + \delta_j + \lambda_t + \epsilon_i \quad (5)$$

which regresses the outcome of interest on whether a student is in the treated cohort, controlling for student baseline characteristics, school fixed effects, and year fixed effects.

4.2.2 Results

This exercise yields similar findings to the difference in differences analysis. Table 4 shows effect sizes of having METCO students in the cohort as close to zero for both Math and English. Columns (1) through (5) show the effects of having METCO peers compared to surrounding cohorts without METCO and Columns (6) through (10) show the opposite effect. Column (3) rules out negative peers effects larger than -0.04 standard deviations for third grade Math and English. Fourth and fifth grade estimates have slightly larger confidence intervals, but can rule out negative effects larger than -0.05 for fourth grade and -0.09 for fifth grade. Attendance and suspension effects are precise zeros, ruling out effect sizes larger than one percentage point in either direction. The effect of having no METCO peers relative to the prior and following cohorts that do have METCO is similarly null, as shown in Columns (5) through (10).

Table A.5 shows similar analysis, but where treatment is defined as one or two years of METCO enrollment that is preceeded and followed by no METCO enrollment, along with the reverse. This alternative definition yields similar findings of no significant peer effects from having METCO students in the cohort.

4.3 Approach 3: Class Size Instrument

4.3.1 Empirics

Teacher union contracts or district rules cap class sizes in each of the districts that accept METCO students. As a result, districts have a financial incentive fill empty seats classroom seats with METCO students, but not exceed the class size cap. Exceeding the cap would require hiring a new teacher, whereas schools face lower per pupil costs by filling empty classroom seats. The 2SLS analysis harnesses the fact that classroom space constraints play a role in determining the number of METCO students a school accepts each year in a method similar to Angrist and Lang (2004)'s adaptation of Maimonides' rule.

District leaders look at their expected enrollment in the Spring, consider class size constraints, and notify METCO of how many students they want to enroll from the city for each grade-level. To instrument for the number of METCO students in a grade, I estimate the predicted class size based on June enrollment in the prior grade using Maimonides' rule (see (Angrist and Lavy, 1999)):

$$r_{gjt} = \frac{e_{gjt}}{int(\frac{e_{gjt}}{c_{gjt}} + 1)} \tag{6}$$

where e_{gjt} reflects resident, non-METCO enrollment in grade g, school j, and year t. The class size limit c_{gjt} is specific to each grade, school, and year and comes from collected data on teacher union contracts and district policies (see Online Appendix). The predicted class size for grade g, school j, and year $t$is represented by r_{gjt}. Under this formula, as schools approach enrollment numbers that are multiples of the class size cap, they add another classroom and predicted average class size shrinks.[27] We see this in the dashed line that represents predicted non-METCO class size in Figure 5. For illustrative purposes we set the class size cap in the figure at 25, which is equal to or larger than the class size cap for the METCO districts.[28] The figure shows that as non-METCO enrollment approaches multiples of 25, average predicted class size sharply drops as a new classroom is added.

The Maimonides rule with a class size cutoff of 25 students closely predicts average class size for first grade enrollment of under 100 students, as seen in Figure 5. The prediction becomes less accurate for the

[27] Angrist and Lang (2004) also use predicted class size and integer indicator variables of predicted class size as instrumental variables in their analysis of the Brookline district's METCO program from 1994 to 2000. In a separate analysis, they use $z_{gjt} = min[max(23 - r_{gjt}, 0), 1]$ as an instrumental variable. This instrument predicts that there will be on average one METCO student per classroom if there grade's average predicted class size is 23 students or less, and none if the predicted class size is larger. This pattern is not visible in this paper's data, or Brookline in the years 2001 - 2020, the years of this study. As such, I do not use this instrument in the analysis.

[28] Except for Scituate which targets a range of 15 - 30 students per class instead of a cap.

larger cohorts with more classes. Because of this, the model described below will control for the predicted number of classes that a school by year cohort has.

Another way to visualize the quality of the prediction is to plot actual average non-METCO class size against average predicted non-METCO class size, shown in Figure 6. Since all of the data in the graph is at the school by year level for first grade, I can use the district-specific class size caps in the predicted class size instead of the generic class size cap of 25 students. Like in the figure above, the predicted class size is close to the actual class size in the majority of cases, with data clustering along the 45 degree line.

In Figure 7, we see that as the average non-METCO class size increases, the average number of METCO students per class falls from about 0.6 students for class sizes of 17 to close to zero students for the average class-size cutoff of 24 students. This observation, combined with the strong relationship between predicted and actual class size, supports the premise for the instrumental variable that as the predicted class size of a school by grade cohort increases, the number of accepted METCO students decline. Therefore, I use predicted class size as an instrument for a school by year METCO enrollment in first grade. To allow for a non-linear relationship between predicted class size and METCO enrollment, I also use indicator variables for integer values of predicted class size as instruments.

The second stage regression estimates the impact of having METCO peers in your cohort on individual outcomes as follows:

$$y_{gjt,i} = \tau a_{gjt,i} + \alpha n_{gjt,i} + \kappa_2 c_{gjt} + \delta_{a2} int.r_{gjt} + \delta_{b2} linear.r_{gjt} + \beta_{\alpha 2jg} + \gamma_{2t} + X_i'\Gamma_2 + \epsilon_{gjt,i} \tag{7}$$

where $y_{gjt,i}$ represents student i's outcome in grade g, school j, year t. The variable $a_{gjt,i}$ is five times the proportion of METCO students in student i's grade, school, and year. Therefore $\tau$is interpreted as the effect of a five percentage point increase in the proportion of METCO students in a grade. For a 20 student class, $\tau$is equivalent to the impact of one additional METCO student.

I control for average non-METCO class size, $n_{gjt,i}$, in a given grade, school, and year. Controls also include indicator variables for total grade enrollment bins, c_{gjt} [29] and the instrumental variables: integers of the predicted class size variable $int.r_{gjt}$ and predicted class size $linear.r_{gjt}$. I include school and year fixed effects, $\beta_{\alpha 2jg}$ and γ_{2t} as well as individual baseline covariates $X_i^{'}$ which include race, ethnicity, and sex indicators, free and reduced price lunch status, special education status, and English language learner status.

The first stage equation is:

$$a_{gjt,i} = \beta_{1jg} + \gamma_{1t} + \delta_{a1} int.r_{gjt} + \delta_{b1} linear.r_{gjt} + \kappa_1 c_{gjt} + X_i'\Gamma_1 + \nu_{gjt,i} \tag{8}$$

$$n_{gjt,i} = \beta_{1jg} + \gamma_{1t} + \delta_{a1} int.r_{gjt} + \delta_{b1} linear.r_{gjt} + \kappa_1 c_{gjt} + X_i'\Gamma_1 + \nu_{gjt,i} \tag{9}$$

Both the proportion of METCO students, a_{gjt}, and the average number of non-METCO students per class n_{gjt}. are endogenous variables: higher performing schools may be pre-disposed to have smaller class sizes and accept more METCO students. Since all of the instruments are non-linear functions of cohort enrollment e_{gjt}, estimates that treat average non-METCO class size as endogenous may be imprecise. However, we can reject the null hypothesis that the model is underidentified with the linear and non-linear predicted class size as instruments (p-value of the Anderson canonical correlation LM statistic is 0.000 for all outcome variables). Therefore, the linear and non-linear predicted class size instruments satisfy the rank condition.

For the predicted class size instruments to meet the monotonicity assumption, we need increases in predicted class size to have either a neutral or negative effect on the average number of METCO students in class for each individual school and first grade cohort. This seems plausible if it didn't hold true that would mean when schools are faced with less space in classrooms for METCO students, they choose to accept more.

The independence assumption requires that fluctuations in the first grade cohort size (the variable input in the predicted class size formula) has a "as good as random" relationship with the relevant observable and unobservable variables. To test the former, I regress each of our baseline covariates on predicted class size and find no statistically significant relationship between demographics and cohort size. This assumption

[29] This controls for the lower predictive power of the instrument for larger enrollment cohorts (referenced above in the discussion of Figure 5) in a non-linear way. The enrollment bins roughly proxy for the number of classrooms in that grade through the formula $c_{gjt} = int(e_{gjt}/25 + 1)$.

also requires that unobservables like racial attitudes of the child or their family, home environment, and motivation are not correlated with cohort size.

The exclusion restriction requires that predicted class size only affects student outcomes through differences in METCO exposure. However, predicted class size and the treatment of increased METCO exposure also effects actual class size which in turn affects student outcomes (see Hanushek (2003) for review of evidence). We can consider the increased class size as a result of METCO a part of the METCO exposure treatment, but to satisfy the exclusion restriction, we need to address the relationship between predicted class size and actual class size. This problem is bounded by the magnitude of the class size changes observed in the data: predicted class sizes range from 17 to 25 students.

While the state administrative data spans from 2001 - 2020, it only has class size data for 2012 - 2020. This presents a challenge in estimating equations 7 for the earlier years without non-METCO class size, n_{gjt}. To adjust for this, I use data from 2012 - 2020 to estimate the following equation:

$$a_{gjt,i} = \beta_{1jg} + \gamma_{1t} + \delta_{a1} int.r_{gjt} + \delta_{b1} linear.r_{gjt} + \kappa_1 c_{gjt} + X_i' \Gamma_1 + \nu_{gjt,i} \tag{10}$$

Then I use those estimated regression coefficients to predict $\hat{n}_{gjt,i}$ for the 2001 - 2011. Finally I estimate the second stage regression, equation 7 on the full sample, but using $\hat{n}_{gjt,i}$ for the earlier years. Results are robust to using just 2012-2020 data, but estimates are more precise with the full sample.

4.3.2 Results

Similar to the other identification strategies, I find no evidence of a significant effect of METCO students on their suburban peers from the two stage least squares estimates. Table 5 shows that the estimates are less precise than the other methods. The first stage for proportion of students in METCO have Angrist and Pischke F-stats of over 100, signaling a valid first stage. For math, I can rule out that a five percentage point increase in METCO students generates effects larger than -0.1 standard deviations on Math and English test scores. For attendance, I can rule out effects larger than a 0.7 percentage point drop for most grades, and 1.7 pecentage point drop for grade 2. Lastly, the two stage least squares estimates can rule out effects above a 1.2 percentage point increase in the suspension rate. In all, I can rule out substantial negative peer effects from adding an additional METCO student to a classroom from the instrumental variable estimates.

5 Heterogeneity

The above analysis conclusively shows no causal impact of having METCO students in a cohort or additional METCO students in a classroom on suburban students' behavioral and test score outcomes. The METCO program may have heterogeneous effects across student characteristics or by the proportion of the classroom, grade, or district that has METCO students.

Having additional urban students of color in the classroom could have a differential effect on students of color who live in the suburbs. Similarly, having additional students from low-income families through the METCO program could have a different effect on suburban students from low income families compared to their relatively wealthier peers. Subgroup analyses for these populations does not yield statistically significant findings for any of the empirical methods, so I cannot conclude whether effects are different for suburban students of color or low-income students (estimates provided at the request of the author). The standard errors on these subgroup analyses are large because suburban Black, Latinx, and lower income students comprise two, three, and six percent of the suburban resident sample respectively (see Table 1).

If effects varied by the proportion of peers from the METCO program, we would expect to see heterogeneous district effects. Figure A.4 shows the two-stage least squares individual district estimates for the subset of estimates with F-statistics above 10. While for Math, English, and attendance have one to three districts with significantly positive or negative estimates, these are not robust across grades and the vast majority of districts have estimates not statistically significantly different from zero.

Within a cohort, some students may have more METCO peers in their classes than others. Schools may group METCO students together so that they are not the only METCO student in their class. In addition, schools may assign METCO students to classes with certain student, teacher, and curricular characteristics. If only a subset of suburban students have METCO students in their classes, then the overall analysis could

mask heterogeneous effects. Classroom-level data shows that on average, 47 percent of the core classes in grade cohorts with METCO students have at least one METCO student. Of the cohorts with METCO students, only five percent of cohorts have METCO students in every core class and 43 percent of suburban resident students have METCO students in at least one class in a given year. Thirty-two percent of resident students with METCO students in their cohort never have a METCO student in one of their core classes in grade K through 12.[30]

Another identification strategy that accounts for how suburban and METCO students are assigned to classes would be needed to estimate differential effects by classroom exposure. Since the classroom assignment processes vary across schools and are often non-random it is difficult to properly estimate the causal impact at the classroom level. Instead, the next section describes which students have more METCO peers in their classes to understand who receives a stronger METCO "treatment." It also details how teacher and classroom characteristics vary in classes with and without METCO students.

6 Class and Teacher Characteristics

6.1 Differences within School and Grade Cohorts

Table 6 compares the traits of core classes with and without METCO students in schools and grade cohorts with METCO. Column 1 shows the average traits of classes without METCO, while Columns 2 and 3 display the difference between classes with and without METCO students, controlling for school, grade, and year fixed effects. Teacher qualifications appear similar among teachers who teach METCO students compared to teachers of classes without METCO students: teacher licensure rates, the state's "highly qualified" status, advanced degree rates, and average years of teaching experience are all similar. Classes with METCO are significantly more likely to be taught by Latinx or Black teachers and less likely to be taught by white or Asian teachers, though the differences are small. For example, 1.9 percent of classes with METCO students are taught by a Black teacher compared to 1.2 percent of classes without METCO students. The average value-added of teachers is 0.51 standard deviations higher in classes with METCO students than those without for Math, but teacher value-added is similar for English.[31]

Classes with METCO students are 5.9 percentage points less likely to be classified as advanced coursework,[32] but also slightly less likely to be classified as remedial. Lower-performing suburban students are assigned to classrooms with METCO students. Suburban students in classes with METCO students scored on average 0.127 standard deviations lower on Math and 0.049 standard deviations lower on English compared to suburban students assigned to classes without METCO students. Classes without METCO students have about a 1.3 standard deviation ability distribution (difference between the 90th and 10th percentile of students' lagged test scores). Students in classes with METCO students experience a significantly wider skill distribution. A small portion of this is due to METCO students having on average lower test scores than suburban resident students. The ability distribution of only the resident students in classes with METCO students shows that on average schools place METCO students in classes with wider ability distributions, even without considering METCO students' test scores.

While there is a wider ability distribution, classes with METCO kids are less likely to have residents with special education or English Language Learner services. There is no significant difference in the presence of low-income students. Classes with METCO students are on average 2.8 students larger than those without.

6.2 Impact of METCO Enrollment on Grade Cohorts Classrooms

Having described the types of students and teachers in classrooms with METCO students, I next use the difference-in-differences method, to explore whether cohorts with METCO students have different classroom

[30] Author's calculations.

[31] Teacher value-added is estimated on suburban resident students using a model with teacher and class-level random effects. Controls include lagged centered test scores, demographics (female, Black, Latinx, free or reduced price lunch), special education, English learner status, and year of birth. All test scores are centered to the state average in that year and grade. Value-added is only available for Math and English teachers in grades 4 - 8 and 10 (when tests are administered and there is a lagged test score). Estimates leave out data from the year they are estimating.

[32] This includes Advanced Placement, International Baccalaureate, courses with "advanced" in the subject name, and those categorized by MA DESE as advanced coursework.

and teacher characteristics than those without. The METCO program may attract teachers to the suburban districts if they are excited about teaching more diverse classes. Alternatively, it may deter teachers if they find teaching classes with METCO students challenging (for example, if there are changes in classroom behavior or the ability distribution). As a result, the METCO program may impact the types of teachers suburban students have.

Columns 4 through 9 of Table 6 show that school participation in METCO has minimal effect on the teacher characteristics a cohort experiences. Teacher qualifications, years of experience, race, and value-added are largely similar when schools start or stop accepting METCO students. The two exceptions are that cohorts after METCO enrollment ended have 3 percentage points fewer new teachers to the school and marginally significantly higher English teacher value-added.

To investigate whether the presence of METCO students in a cohort affects the distribution of ability levels in a class, I look at the average difference between the 90th and 10th percentiles' lagged test scores in students' average core classes. If enrolling METCO students leads to more tracking, we would expect a narrowing in the lagged test score distribution of classes and a negative point estimate for when METCO programs start (Column 5). When including METCO students, the 90 - 10 lagged test score spread is 0.085 standard deviations higher in Math and similar in English when METCO enrollment starts. However, the 90 - 10 spread of just resident students is not statistically significantly different both when METCO enrollment starts and ends. This means there is not evidence of METCO changing the ability distribution in classes outside of relatively small widening of the Math distribution. It also shows that schools do not reduce or increase tracking in response to METCO students and that the middle 80 percent of the class has a similar range of abilities before and after METCO enrollment.

METCO cohorts have on average 0.72 more METCO students per 20 student class and 1 percentage point fewer suburban special education students. The representation of economically disadvantaged students and English Language Learners are similar across cohorts with and without METCO. Lastly, METCO enrollment reduces class size by about half a student when schools switch to accepting METCO students and increases average lass size by about 0.35 students when districts stop accepting METCO students.

Therefore, the presence of METCO students in a cohort does not significantly change the classroom or teacher characteristics beyond the demographic changes from having the METCO students and slightly decreasing class size.

7 Conclusion

Efforts to increase racial and socioeconomic diversity in schools and neighborhoods through school assignment or housing policy are often met with concern about negative peer effects. Two possible channels for negative peer effects could be increased class size and widening skill distribution within a classroom. Parents may be concerned that these two factors could reduce the attention their students receive and worsen the match between the classroom content and their students' ability level.

Difference-in-differences analysis shows that after enrolling METCO students, schools shrink class sizes by about 0.5 students. This could be due to schools starting to accept METCO students when projected enrollment is falling. Conversely, class sizes slightly increase when METCO enrollment ends in a school, possibly in anticipation of increasing cohort sizes.

The ability distribution in classes remains similar. The ability distribution widens by four percent in Math and remains similar for English. Furthermore, schools do not increase or decrease tracking as a result of changes in METCO enrollment policy. Therefore, neither the class size channel nor the ability distribution channels are viable in this case.

In addition to ruling out two potential channels for negative peer effects, I use three different identification strategies to find no evidence of substantial negative test score or behavioral peer effects of enrolling up to 10 percent of a cohort with urban students of color. Effects are consistently null across districts and cohorts with relatively more and fewer urban students. The three methods each have different interpretations. An instrumental variables approach based on class size constraints finds no effect from increasing the proportion of urban students in a suburban class. The fixed effects approach finds no difference among cohorts with and without METCO students when districts sporadically change whether they enroll urban students from year to year. Lastly, the difference-in-differences approaches finds no difference in test score and behavioral

outcomes when districts start or stop accepting METCO students. The conclusive null effect with three different local average treatment effects suggests that concerns of negative academic and behavioral effects of urban students of color in the suburban classrooms are not substantiated by data.

With increasing racial and economic segregation in American public schools, it is important to understand the peer effects of increasing integration. Continued work will incorporate older cohorts to look at longer run outcomes, including college enrollment and graduation, employment and earnings, civic engagement, criminal justice outcomes, marriage and fertility and housing outcomes. It is of particular interest whether increased integration changes the social outcomes and political beliefs of suburban students.

References

Angrist, Joshua, and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Student Achievement." *Quarterly Journal of Economics*, 114(2): 533–575.

Angrist, Joshua D, and Kevin Lang. 2004. "Does school integration generate peer effects? Evidence from Boston's Metco Program." *American Economic Review*, 94(5): 1613–1634.

Armor, David J. 1972. "The Evidence on Busing. Research Report." *Public Interest*, 28: 90–126.

Baker, Andrew, David F. Larcker, and Charles C. Y. Wang. 2021. "How Much Should We Trust Staggered Difference-In-Differences Estimates?" *European Corporate Governance Institute – Finance Working Paper*, 736.

Bergman, Peter, et al. 2018. "The Risks and Benefits of School Integration for Participating Students: Evidence from a Randomized Desegregation Program." Institute for the Study of Labor (IZA).

Boardman, Richard, and Linda Brandt. 1968. "Metco - a Descriptive Report." *Education Resource Information Center*, ERIC No: ED088225.

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2020. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics.*

Card, David, and Jesse Rothstein. 2007. "Racial segregation and the black–white test score gap." *Journal of Public Economics*, 91(11-12): 2158–2184.

Cengiz, Doruk, Arindrajit Dube, Attila Linder, and Ben Zipperer. 2019. "The Effect of Minimum Wages on Low-Wage Jobs." *The Quarterly Journal of Economics*, 134(3): 1405–1454.

Clarke, Keith W. 1975. "A Descriptive Study of the Metco Program." *Education Resource Information Center*, ERIC No: ED124631.

Deshpande, Manasi, and Yue Li. 2019. "Who Is Screened Out? Application Costs and the Targeting of Disability Programs." *American Economic Journal: Economic Policy*, 11(4): 213–248.

Eaton, Susan. 2001. *The Other Boston Busing Story.* New Haven, CT:Yale University Press.

Elliott, Cary A. 1998. "Get on the Bus? The Long Run Effect of Metco Suburban Education on Inner-City Students." *Chapter II in Three Essays in Applied Microeconomics, Princeton University Economics Department, Ph. D. thesis, November.*

Figlio, David, Paola Giuliano, Riccardo Marchingiglio, Umut Özek, and Paola Sapienza. 2021. "Diversity in Schools: Immigrants and the Educational Performance of U.S. Born Students." *NBER Working Paper.*

Godsil, Rachel D, Linda R Tropp, Phillip Atiba Goff, and John A Powell. 2014. "Addressing implicit bias, racial anxiety, and stereotype threat in education and health care." *The Science of Equality*, 1.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*.

Government Accountability Office. 2016. "K-12 Education: Better Use of Information Could Help Agencies Identify Disparities and Address Racial Discrimination." *Report GAO-16-345*, Washington, D.C.: U.S. Government Printing Office.

Guryan, Jonathan. 2004. "Desegregation and Black Dropout Rates." *The American Economic Review*, 94(4): 919–943.

Hanushek, Eric. 2003. "The Failure of Input-based Schooling Policies." *The Economic Journal*, 113: F64–F98.

Johnson, Rucker C. 2011. "Long-run impacts of school desegregation & school quality on adult attainments." National Bureau of Economic Research.

Lazear, Edward P. 2001. "Educational production." *The Quarterly Journal of Economics*, 116(3): 777–803.

Mantil, Ann. 2022. "Crossing District Lines: The Impact of Urban-Suburban Desegregation Programs on Educational Attainment." *Educational Evaluation and Policy Analysis*, 44(1): 127–148.

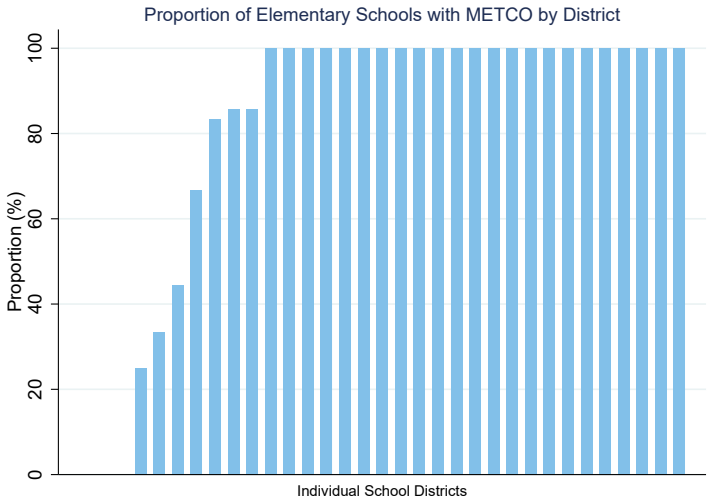
Orfield, Gary, Jennifer Arenson, Tara Jackson, Christine Bohrer, Dawn Gavin, and Emily Kalejs. 1997. "City-Suburban Desegregation. Parent and Student Perspectives in Metropolitan Boston."

Sun, Liyang, and Sarah Abraham. 2020. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*.

Vigdor, Jacob, and Jens Ludwig. 2007. "Segregation and the black-white test score gap." National Bureau of Economic Research.

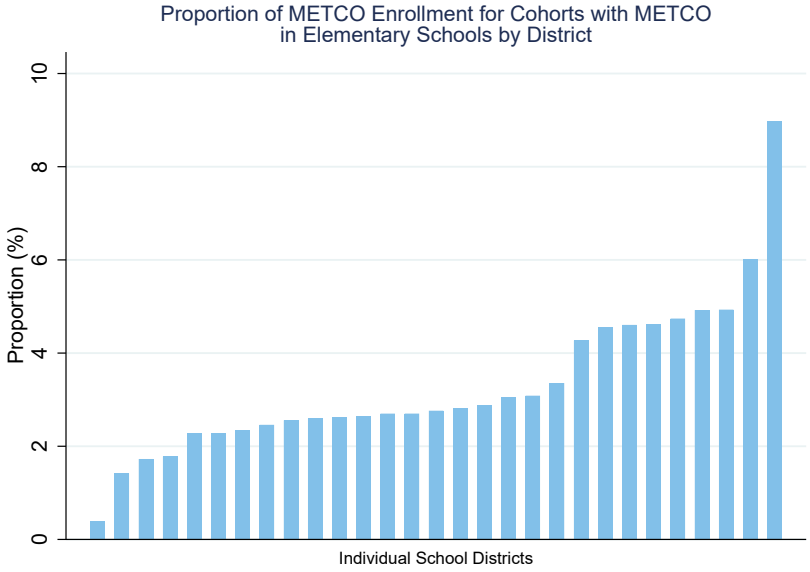
Zebrowitz, Leslie A, Benjamin White, and Kristin Wieneke. 2008. "Mere exposure and racial prejudice: Exposure to other-race faces increases liking for strangers of that race." *Social cognition*, 26(3): 259–275.

Figure 1: Proportion of Elementary Schools with METCO by District



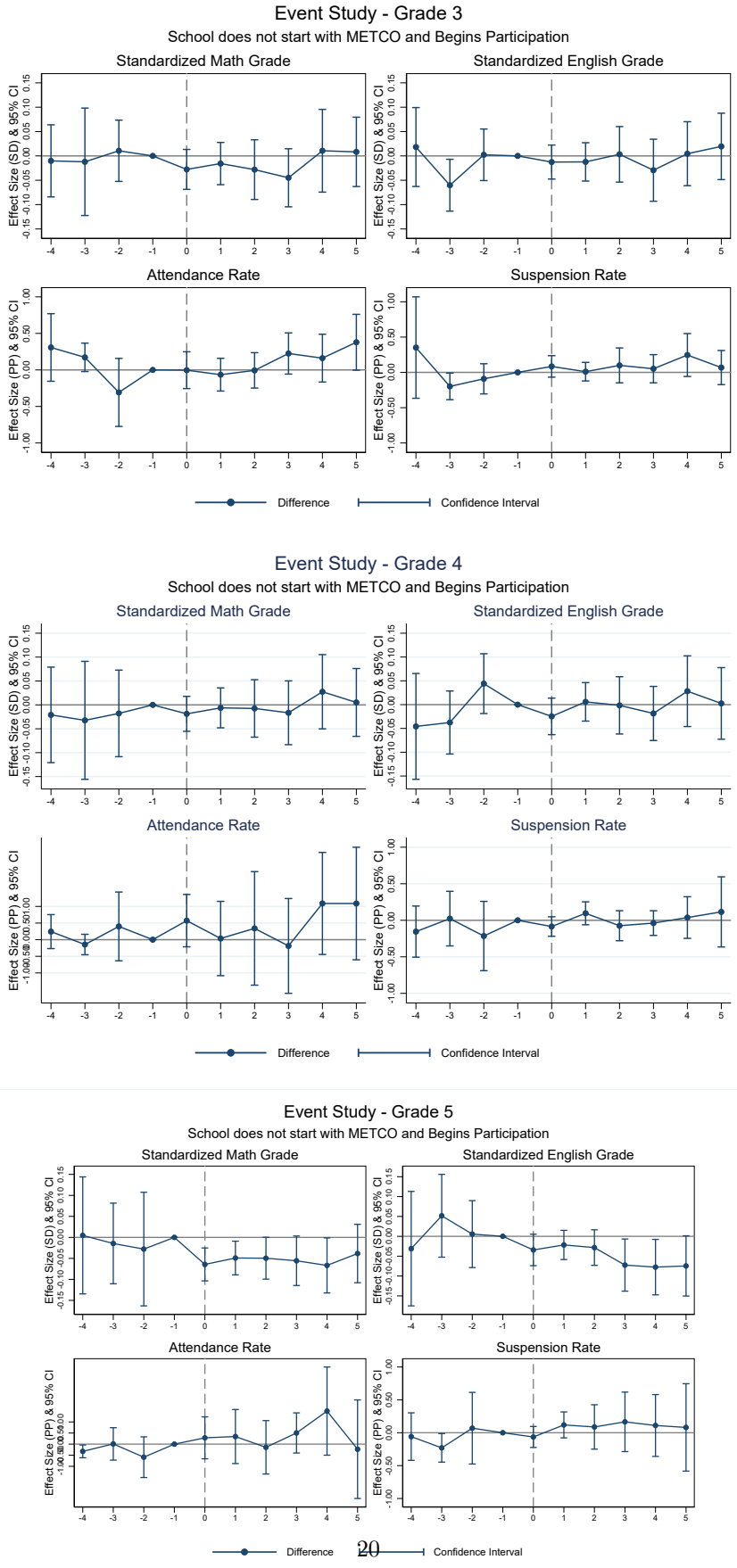
Notes: Each bar represents an individual school district and the proportion of its elementary schools that enroll METCO students. The figure only displays suburban districts that accept METCO students and have elementary schools.

Figure 2: Proportion of METCO Students in Treated Cohorts



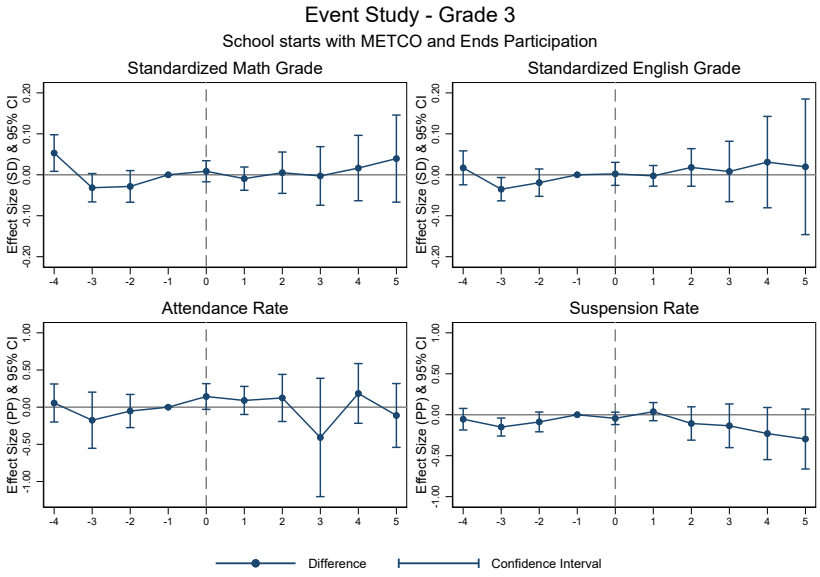
Notes: Each bar represents an individual school district and plots the average proportion METCO students in grade by school cohorts that have any METCO students. The figure only displays suburban districts that accept METCO students and have elementary schools.

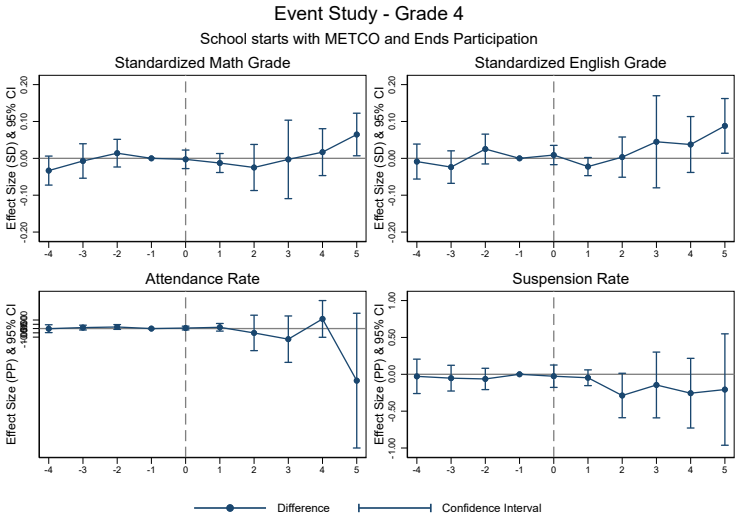
Figure 3: Event Study Results: Begins METCO Participation

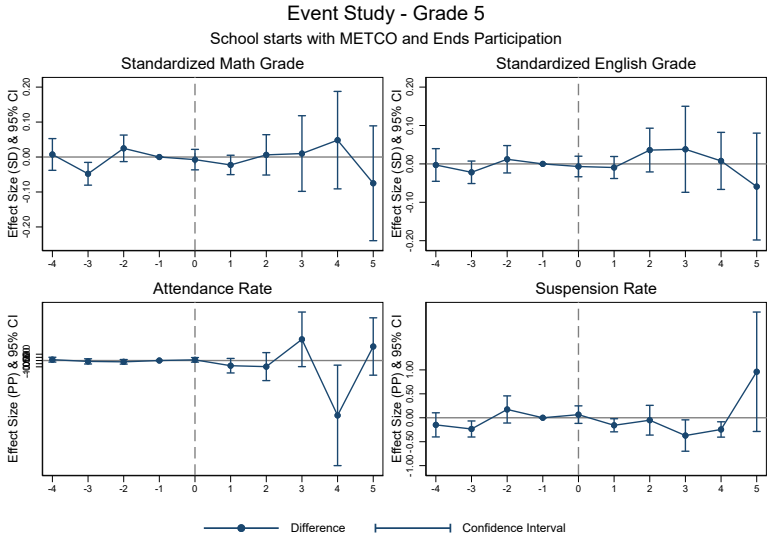


Notes: This figure shows the event study estimates of Equation 4 for grades 3 - 5 outcomes for schools that began their METCO participation.

Figure 4: Event Study Results: Ends METCO Participation

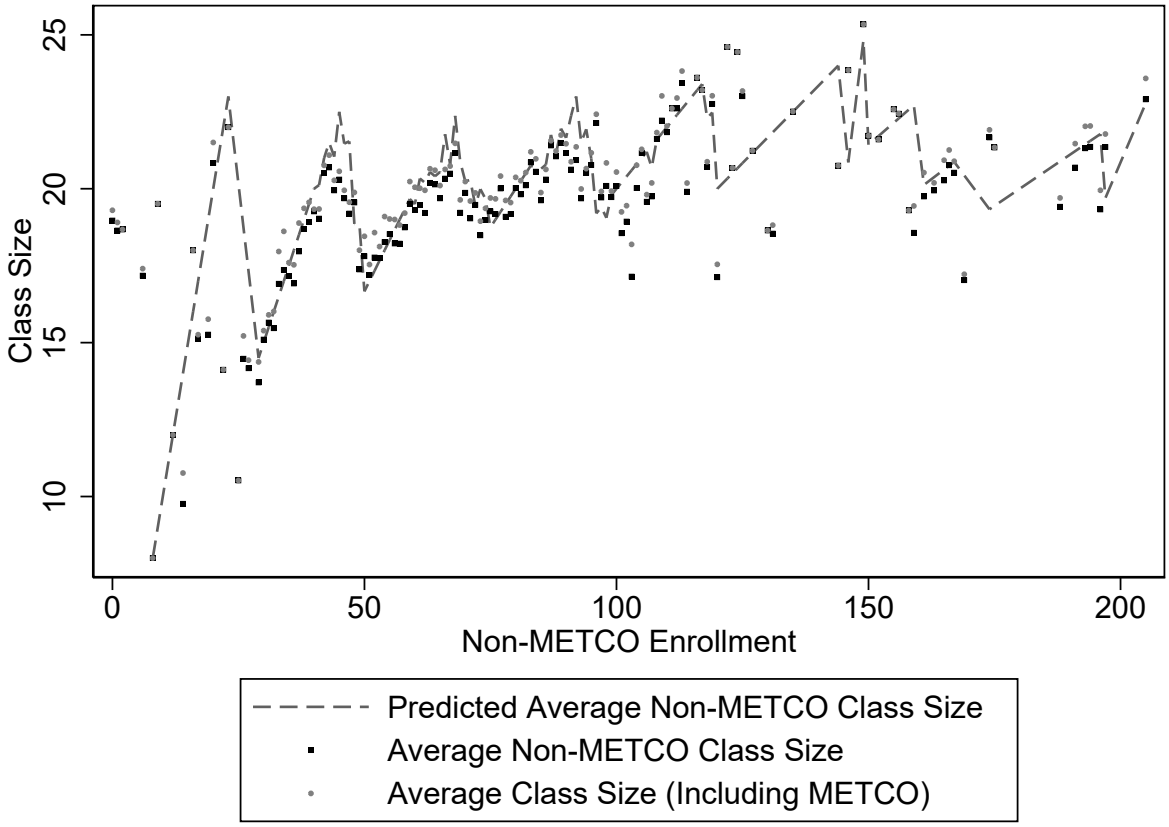






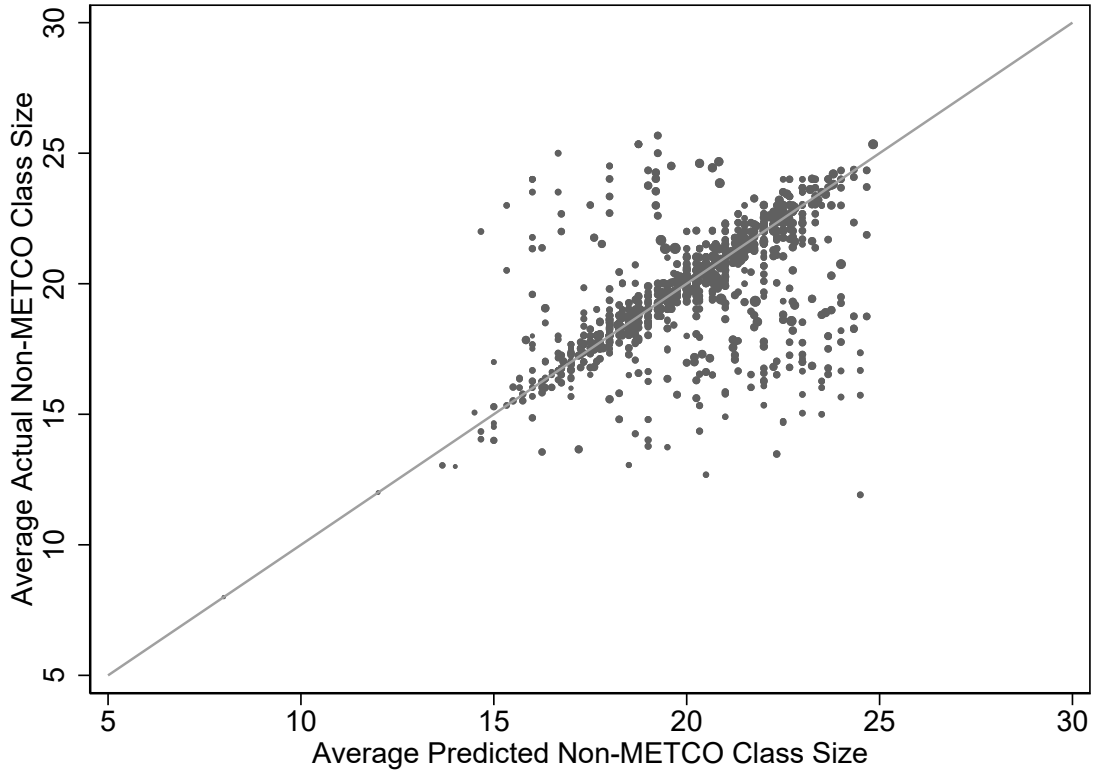
Notes: This figure shows the event study estimates of Equation 4 for grades 3 - 5 outcomes for schools that ended or paused their METCO participation.

Figure 5: Predicted class size and actual class size in first grade



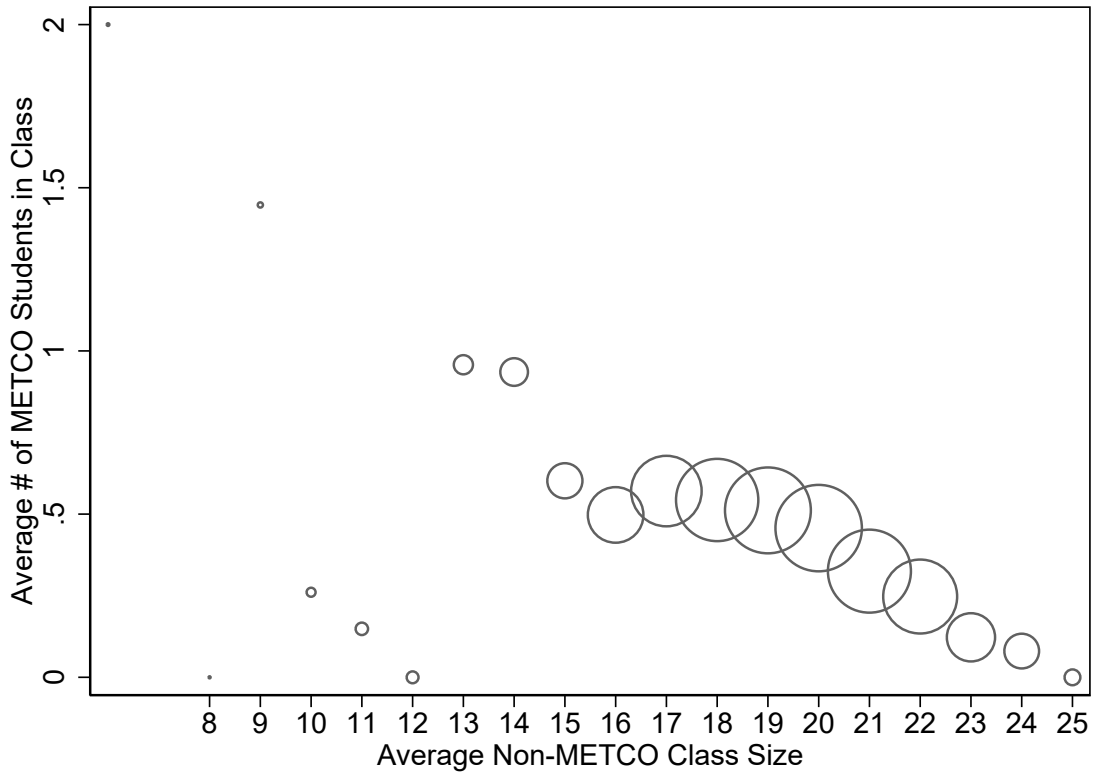
Notes: Each data point reflects an individual school by year for first grade students. Predicted Non-METCO class size is estimated using Angrist and Lavy (1999)'s Maimonides rule with a class size cutoff of 25 students. Actual class size cutoffs in METCO suburban districts range from 20 to 30 students with an average of 24 students per class.

Figure 6: Relationship between predicted and actual class size



Notes: This figure plots a 45 degree line and the average non-METCO class size in suburban METCO receiving districts against the predicted non-METCO class size according to the Maimonides Rule described in equation 6. Data is for first grades and at the school by year level. The predicted class size calculation uses district-specific class size cutoffs for first grade.

Figure 7: Relationship between number of METCO students in class and average non-METCO class size



Notes: This figure plots average number of METCO students per class against binned, integer values of average non-METCO class size for the first grade. Circle size is average weighted by first grade enrollment with larger circles signifying more common average non-METCO class sizes.

Table 1: Descriptive Statistics of K-12 Data

	METCO Suburban Districts	METCO Urban Students	Boston & Springfield Public Schools	Non-METCO Districts in Metro-area
	(1)	(2)	(3)	(4)
Female	0.49	0.55	0.48	0.49
Black	0.02	0.79	0.37	0.08
Latino/a	0.03	0.16	0.41	0.11
Asian	0.08	0.04	0.07	0.05
White	0.86	0.02	0.15	0.75
Subsidized Lunch	0.06	0.51	0.79	0.24
Special Education	0.13	0.11	0.16	0.14
Immigrant	0.03	0.02	0.12	0.05
English-Language Learner	0.04	0.05	0.28	0.09
Ever Suspended	0.02	0.08	0.11	0.07
Attendance Rate	0.95	0.94	0.87	0.93
English Exam	0.49	-0.13	-0.63	0.08
Math Exam	0.51	-0.28	-0.58	0.07
Observations with Test Scores	718,758	20,814	418,884	1,943,528
Observations	2,582,019	68,289	1,651,745	6,970,946
Unique Students	417,314	11,438	315,621	1,132,711
Years of Data	21	21	21	21
Unique Schools	247	211	239	702
Unique Districts	38	38	2	101

Notes: This table shows descriptive statistics for students in suburban districts that receive METCO students, Boston and Springfield students that participate in METCO, the districts that send students to the METCO program (Boston and Springfield public schools), and non-METCO suburban districts in the Boston and Springfield metro area. Data is at the student by year level. Test scores are centered at the state mean for each grade and year. State exams are given in grades 3 - 8 and 10. Characteristics in rows 1-8 (female through English Language Learner) are baseline characteristics.

Table 2: Ordinary Least Squares Estimates of the Relationship between METCO and Outcomes

	Metro-Area			Receiving Districts				
	Non-METCO mean	Relationship between METCO and Outcome		Non-METCO mean	Relationship between METCO and Outcome			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Math	0.091	0.414	0.015	0.295	0.154	0.014	0.004	0.007
		(0.002)	(0.003)		(0.003)	(0.003)	(0.003)	(0.003)
N		4,004,646	4,004,646		1,147,405	1,147,405	1,004,827	977,039
English	0.098	0.371	0.013	0.307	0.128	0.012	0.006	0.009
		(0.002)	(0.003)		(0.002)	(0.003)	(0.003)	(0.003)
N		3,989,762	3,989,762		1,142,736	1,142,736	1,000,257	977,547
Attendance Rate	0.867	0.015	0.004	0.881	0.004	0.004	0.002	0.002
		(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)
N		9,090,036	9,090,036		2,575,712	2,575,712	2,159,475	1,943,228
Suspended	0.042	-0.027	-0.004	0.014	-0.006	0.001	0.001	0.001
		(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)
N		8,625,992	8,625,992		2,452,107	2,452,107	2,159,475	1,943,228
Number of Suspensions	0.134	-0.096	-0.017	0.038	-0.021	0.001	0.002	0.002
		(0.001)	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
N		8,625,992	8,625,992		2,452,107	2,452,107	2,159,475	1,943,228
Year & Grade FE, Region Control		X	X		X	X	X	X
District FE			X			X	X	X
Lagged School Traits							X	X
Individual Baseline Controls								X

Notes: This table reports the OLS estimates of the relationship between the proportion of METCO students in a schoolXgrade and the outcomes of non-METCO students in that suburban district. The endogenous variable is the ratio of METCO to non-METCO students in a schoolXgrade multiplied by 20 so that a one unit increase in the endogenous variable can be interpreted as adding one METCO student to a 20-person class. All models control for whether the district is in the Springfield metropolitan area and include year and grade fixed effects. Lagged school traits include the prior year's average attendance rate, suspension rate, days suspended, and standardized Math and English test scores as well as the proportion of students receiving free or reduced price lunch, special education or English Language Learner services, and racial demographics. Individual baseline controls include the following information for the first year students appear in the data: gender, race and ethnicity, free and reduced price lunch, special education and level of services, English Language Learner and level of proficiency, and immigrant.

Table 3: Difference-in-Differences Results

	School does not start with METCO and Begins Participation					School starts with METCO and Ends Participation				
	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Math			0.002	0.005	-0.004			-0.003	-0.008	-0.016
			(0.015)	(0.018)	(0.016)			(0.014)	(0.013)	(0.015)
Observations (Students)			990,879	930,671	481,247			1,174,017	1,081,556	787,348
Treatment Groups			143	144	128			156	163	146
Control Schools			4,653	4,690	3,896			5,856	6,107	5,251
English			0.006	0.025	-0.013			0.008	-0.016	-0.004
			(0.015)	(0.015)	(0.016)			(0.013)	(0.014)	(0.015)
Observations (Students)			1,104,548	932,729	481,977			1,283,026	1,073,060	788,900
Treatment Groups			155	144	128			168	163	146
Control Schools			5,101	4,690	3,899			6,333	6,107	5,251
Attendance Rate	-0.001	0.000	0.000	-0.002	0.002	0.003	-0.001	0.000	-0.002	-0.009
	(0.001)	(0.001)	(0.001)	(0.004)	(0.005)	(0.001)	(0.001)	(0.001)	(0.003)	(0.005)
Observations (Students)	1,352,522	1,186,758	1,126,408	1,025,589	545,921	1,719,896	1,621,151	1,306,911	1,137,860	875,253
Treatment Groups	156	156	155	155	140	169	169	168	168	157
Control Schools	5,176	5,137	5,103	5,040	4,293	6,374	6,374	6,337	6,289	5,632
Suspended	0.001	0.001	0.000	0.001	0.001	0.000	0.000	0.000	-0.001	-0.001
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
Observations (Students)	1,200,955	1,129,711	1,126,337	1,025,567	545,921	1,571,442	1,558,545	1,306,906	1,137,859	875,253
Treatment Groups	142	152	155	155	140	154	164	168	168	157
Control Schools	4,659	4,996	5,103	5,040	4,293	5,791	6,181	6,337	6,289	5,632
Number of Suspensions	0.002	0.002	-0.001	0.001	0.002	-0.002	-0.001	0.002	0.001	-0.002
	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.002	0.001
Observations (Students)	1,200,955	1,129,711	1,126,337	1,025,567	545,921	1,571,442	1,558,545	1,306,906	1,137,859	875,253
Treatment Groups	142	152	155	155	140	154	164	168	168	157
Control Schools	4,659	4,996	5,103	5,040	4,293	5,791	6,181	6,337	6,289	5,632

Notes: This table shows the stacked difference in differences estimates of the impact of schools switching to enrolling METCO students (Columns 1 – 5) and schools switching to not enrolling METCO students (Columns 6 – 10). The stacked difference in differences calculates a weighted average of the individual difference in differences estimates of each school that changed METCO enrollment policy. Treatment schools in Columns 1-5 have a minimum of one year without METCO enrollment in the first grade followed by a minimum of one year with first grade METCO enrollment. Data includes up to four years before the school switched to enrolling METCO first graders and up to four years after. The comparison schools did not accept METCO students in first grade for all the years the treatment school appears in the data. Each individual treatment group and their comparison schools form a policy change subsample. The treatment and comparison schools are chosen using the same rules for the endogenous variable of switching to no longer enrolling METCO students in Columns 6–10. Observations can appear in multiple comparison groups for treatment schools. Treatment observations can appear as comparison observations for other treatment schools. Standard errors are clustered at the individual student level and by the policy change subsample. All estimates include school fixed effects and individual student demographics (gender, race, immigrant status). The sample excludes METCO participants.

Table 4: Sporadic METCO Participation Results

	Effect of Being in a METCO Cohort					Effect of Being in No METCO Cohort				
	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Math			0.022	0.026	0.005			-0.022	0.001	-0.028
			(0.020)	(0.023)	(0.021)			(0.016)	(0.015)	(0.020)
Observations (Students)			7,126	7,061	5,644			15,724	14,482	10,735
Treatment Groups			39	40	34			88	85	69
Control Schools			78	80	68			176	170	138
English			0.032	0.032	0.000			-0.011	-0.029	-0.022
			(0.022)	(0.022)	(0.026)			(0.015)	(0.015)	(0.020)
Observations (Students)			8,125	7,093	5,647			16,554	14,488	10,708
Treatment Groups			44	40	34			92	85	69
Control Schools			88	80	68			184	170	138
Attendance Rate	0.000	0.002	-0.003	0.002	0.016	0.004	0.000	-0.002	0.003	-0.010
	(0.002)	(0.001)	(0.002)	(0.001)	(0.009)	(0.002)	(0.001)	(0.001)	(0.002)	(0.008)
Observations (Students)	8,906	8,541	8,264	7,794	6,671	19,419	18,525	16,808	16,207	13,603
Treatment Groups	44	44	44	43	39	94	94	92	92	85
Control Schools	88	88	88	86	78	188	188	184	184	170
Suspended	0.001	0.000	-0.001	0.003	0.002	-0.001	-0.001	0.001	-0.002	-0.002
	(0.001)	(0.001)	(0.000)	(0.002)	(0.002)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Observations (Students)	7,637	8,255	8,264	7,794	6,671	18,207	18,215	16,808	16,207	13,603
Treatment Groups	38	43	44	43	39	88	93	92	92	85
Control Schools	76	86	88	86	78	176	186	184	184	170
Number of Suspensions	0.004	-0.002	-0.002	0.006	0.002	-0.002	-0.002	0.003	-0.003	-0.003
	(0.003)	(0.002)	(0.002)	(0.005)	(0.003)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)
Observations (Students)	7,637	8,255	8,264	7,794	6,671	18,207	18,215	16,808	16,207	13,603
Treatment Groups	38	43	44	43	39	88	93	92	92	85
Control Schools	76	86	88	86	78	176	186	184	184	170

Notes: This table shows the difference between cohorts with METCO relative to those without METCO after controlling for individual student baseline characteristics (gender, race, immigrant status), school fixed effects, and year fixed effects. The sample for Columns (1) through (5) is restricted to individual students in cohorts that have METCO and the prior and subsequent cohorts do not have METCO. Columns (6) through (10) contain estimates for cohorts that did not have METCO and the cohorts one year before and after did have METCO.

Table 5: Two-Stage Least Squares Results

	3rd Grade Mean	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
	(1)	(2)	(3)	(4)	(5)	(6)
Math	0.440			-0.041	0.035	-0.013
				(0.035)	(0.046)	(0.031)
N				94,964	94,568	94,189
F-stat				228	147	344
English	0.452			0.021	0.068	-0.047
				(0.035)	(0.044)	(0.029)
N				94,964	94,568	94,189
F-stat				228	147	344
Attendance Rate	0.967	0.003	-0.010	-0.002	0.001	0.003
		(0.004)	(0.002)	(0.002)	(0.002)	(0.001)
N		96,513	95,421	94,979	94,572	94,194
F-stat		133	143	229	147	344
Suspended	0.002	-0.004	-0.001	-0.001	0.002	-0.003
		(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
N		96,392	95,421	94,979	94,572	94,194
F-stat		133	143	229	147	344
Number of Discreet Class Sizes		11	11	11	11	11
Number of Cohorts (SchoolXYear)		1625	1681	1679	1679	1671
Number of Schools		144	147	147	147	147
Number of Districts		35	35	35	35	35

Notes: This table shows the two-stage least squares estimates of the impact of an additional METCO student on suburban resident student outcomes. The main endogenous variable is 20 times the average ratio of METCO to non-METCO students in an individual's grade cohort from grade 1 through the outcome grade. A one unit increase in the main endogenous variable represents an additional METCO student in a 20-student classroom. The model includes another endogenous variable: average class size excluding METCO students. This is estimated for 2001-2011 before administrative data on class size exists (see Equation 10). Linear and integer forms of predicted class size (see Equation 6 for estimation strategy) instrument for both the ratio of METCO to non-METCO students and non-METCO class size. Controls include binned total grade-level enrollment that proxy for the number of classrooms in a grade, school and year fixed effects, individual baseline covariates (including gender, race, free and reduced price lunch status, special education status, and English Language Learner status). The sample includes students with non-missing values for the third through fifth grade Math and English exams.

Table 6: Difference in Differences Results for Class and Teacher Traits

				School does not start with METCO and Begins Participation			School starts with METCO and Ends Participation		
	Classes without METCO	Difference from classes with METCO	Standard Error	Cohorts with No	Coefficient	Standard Error	Cohorts with No METCO	Coefficient	Standard Error
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Teacher Traits									
% of Teachers licensed in teaching assignment	0.997	0.001	(0.000)	0.997	0.001	(0.002)	0.998	0.002	(0.002)
% of core academic classes taught by highly qualified teachers	0.988	0.006	(0.001)	0.998	0.005	(0.006)	0.994	0.004	(0.003)
% of teachers with advanced degree	0.027	0.002	(0.001)	0.020	-0.004	(0.012)	0.025	-0.007	(0.011)
Average years of teaching experience in MA	11.114	0.126	(0.032)	10.054	0.138	(0.342)	11.171	-0.369	(0.374)
% novice teacher (<2 years)	0.095	-0.002	(0.001)	0.133	0.015	(0.022)	0.098	-0.014	(0.020)
% new to school (<2 years in school)	0.059	-0.004	(0.001)	0.062	0.002	(0.016)	0.067	-0.033**	(0.016)
% White teachers	0.966	-0.002	(0.001)	0.996	0.004	(0.008)	0.941	0.001	(0.006)
% Latino teachers	0.012	0.003	(0.001)	0.005	0.003	(0.003)	0.027	-0.004**	(0.002)
% Black teachers	0.012	0.007	(0.001)	0.006	0.002	(0.003)	0.031	-0.001	(0.004)
% Asian teachers	0.029	-0.003	(0.001)	0.003	-0.005	(0.008)	0.051	0.007	(0.006)
Average VAM of Math teachers	0.056	0.051	(0.015)	0.374	0.102	(0.076)	0.313	0.007	(0.046)
Average VAM of English teachers	0.199	-0.009	(0.015)	0.139	0.026	(0.057)	0.215	0.078*	(0.043)
% advanced course	0.233	-0.059	(0.001)	0.000	-	(0.000)	0.000	-	(0.000)
% remedial course	0.111	-0.022	(0.001)	0.002	0.010	(0.011)	0.102	-0.037**	(0.015)
Panel B: Class Traits									
Average lagged test score of residents:									
Math	0.349	-0.127	(0.004)	0.466	0.004	(0.022)	0.557	-0.011	(0.018)
English	0.314	-0.049	(0.003)	0.503	0.020	(0.017)	0.528	-0.030	(0.019)
Class 90th - 10th percentile (all students)									
Math	1.303	0.420	(0.003)	1.993	0.085**	(0.037)	2.095	0.002	(0.033)
English	1.270	0.360	(0.003)	1.924	0.014	(0.035)	1.970	0.017	(0.034)
Class 90th - 10th percentile (residents)									
Math	1.303	0.360	(0.003)	1.992	0.037	(0.034)	2.034	0.050	(0.031)
English	1.270	0.309	(0.003)	1.922	-0.010	(0.035)	1.930	0.032	(0.033)
Average # of METCO students	0.000	1.486	(0.002)	0.048	0.720***	(0.045)	0.942	-0.698***	(0.042)
Average % special ed of residents in class	0.288	-0.017	(0.001)	0.192	-0.009*	(0.005)	0.184	0.012***	(0.005)
Average % free-reduced lunch of residents in class	0.117	0.001	(0.001)	0.107	-0.004	(0.004)	0.069	0.000	(0.004)
Average % English Learner of residents in class	0.057	-0.033	(0.001)	0.027	0.004	(0.003)	0.049	-0.002	(0.002)
Average class size	15.483	2.804	(0.034)	21.014	-0.499**	(0.230)	21.406	0.346*	(0.197)

Notes: Column 1 shows the average classroom traits for core classes without METCO students using class by year level data. The results of the regression of each classroom trait on whether a class has METCO students or not, controlling for school, year, and grade fixed effects, appear in Column 2 and 3. They show the within school, grade, and year differences of classes with and without METCO students for grade cohorts with classes both with and without METCO students. Columns 4 – 6 show stacked difference in differences estimates of the impact of schools switching to enrolling METCO students on fifth grade teacher and classroom traits. Columns 7 -9 show the estimates for schools switching to not enrolling METCO students. The stacked difference in differences calculates a weighted average of the individual difference in differences estimates of each school that changed METCO enrollment policy. See Table 3 for regression specifications for Cols 4 – 9. Unless otherwise noted, class averages exclude METCO participants. Only core academic classes are included.

Table 7: Two-Stage Least Squares High School Outcomes Results

	Non-METCO mean (1)	2SLS Estimate (2)	Standard Error (3)	N (4)	F-stat (5)
Panel A: Testing					
Meets Standardized Testing High School Graduation Requirement	0.929	0.001	(0.006)	65142	866
Qualify for Adams Scholarship	0.291	0.003	(0.011)	65216	866
Take SAT	0.652	0.032	(0.010)	75924	1014
SAT 800 or Higher	0.635	0.035	(0.010)	75924	1014
SAT 1000 or Higher	0.534	0.032	(0.011)	75924	1014
SAT 1200 or Higher	0.308	0.008	(0.010)	75924	1014
SAT 1400 or Higher	0.096	0.015	(0.007)	75924	1014
Took AP	0.467	0.003	(0.011)	75924	1014
Number of APs	1.549	0.082	(0.048)	75924	1014
AP Score Above 3	0.424	0.000	(0.011)	75924	1014
AP Score 4 or 5	0.362	0.005	(0.011)	75924	1014
Panel B: High School Graduation					
Dropout	0.005	0.001	(0.002)	73361	907
Graduate in 4 Years	0.858	-0.001	(0.005)	73361	907
Graduate in 5 Years	0.990	0.003	(0.002)	64038	861
Panel C: Post High School Aspirations					
Any College	0.904	0.005	(0.007)	64557	873
2 Year College	0.062	0.003	(0.006)	64557	873
4 Year College	0.843	0.002	(0.009)	64557	873
Number of Discreet Class Sizes		11			
Number of Cohorts (SchoolXYear)		1198			
Number of Schools		140			
Number of Districts		34			

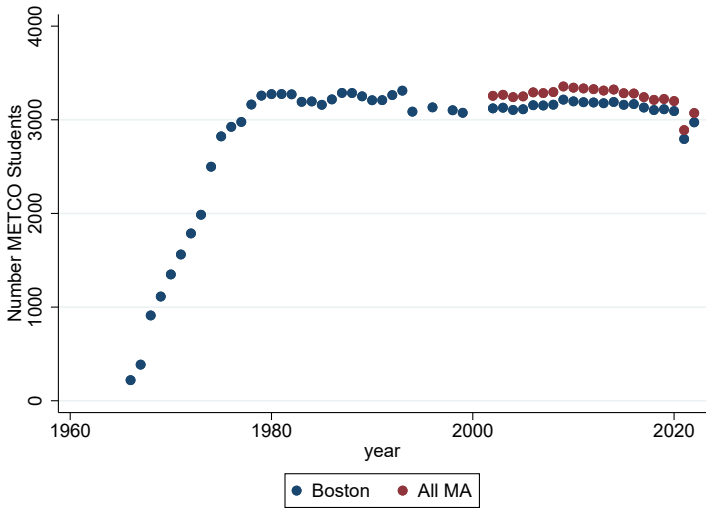
Notes: This table shows the two-stage least squares estimates of the impact of an additional METCO student on suburban resident student high school outcomes. The main endogenous variable is 20 times the average ratio of METCO to non-METCO students in an individual's grade cohort from grade 1. A one unit increase in the main endogenous variable represents an additional METCO student in a 20-student classroom. The model includes another endogenous variable: average class size excluding METCO students. This is estimated for 2001-2011 before administrative data on class size exists (see Equation 10). Linear and integer forms of predicted class size (see Equation 6 for estimation strategy) instrument for both the ratio of METCO to non-METCO students and non-METCO class size. Controls include binned total grade-level enrollment that proxy for the number of classrooms in a grade, school and year fixed effects, individual baseline covariates (including gender, race, free and reduced price lunch status, special education status, and English Language Learner status). Students have a value of "0" for SAT and AP outcomes if they did not take the respective exams. Post-high school plans data come from a survey administered to 10th graders.

Table 8: Two-Stage Least Squares College Outcomes Results

	Non-METCO mean	2SLS Estimate	Standard Error	N	F-stat
	(1)	(2)	(3)	(4)	(5)
Panel A: Racial Composition of College					
Percent Black or Latinx	0.161	-0.002	(0.002)	55641	778
Percent Non-white	0.382	-0.003	(0.004)		
Panel B: College Enrollment					
Any College	0.819	0.011	(0.009)		
Two-Year College	0.101	-0.007	(0.007)		
Four-year College	0.763	0.019	(0.010)	67371	917
Four-year Public	0.318	-0.002	(0.011)		
Four-year Private	0.476	0.023	(0.012)		
Four-year Public in MA	0.170	-0.003	(0.009)		
Panel C: Four-Year College Ranking					
Most Competitive	0.139	0.010	(0.008)		
Highly Competitive	0.278	0.015	(0.010)	67371	917
Very Competitive	0.413	0.006	(0.011)		
Competitive	0.728	0.012	(0.011)		
Panel D: Four-Year College Persistence					
One academic semester	0.757	0.015	(0.010)	67371	917
Three academic semesters	0.730	0.001	(0.011)	58810	954
Five academic semesters	0.705	0.010	(0.012)	50068	840
Seven academic semesters	0.671	0.016	(0.015)	41476	569
Panel E: College Graduation within 6 years					
Any	0.652	0.027	(0.019)		
Two-Year College	0.106	-0.008	(0.012)	32892	423
Four-year College	0.644	0.028	(0.019)		

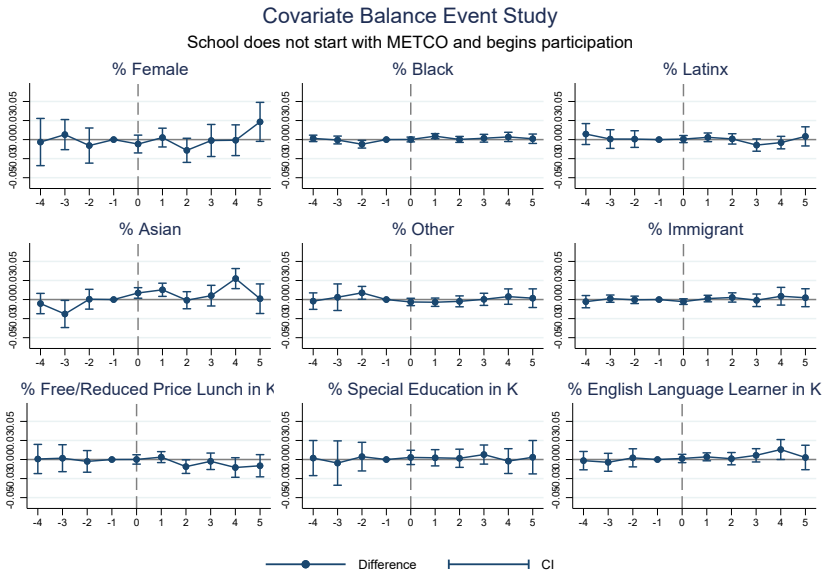
Notes: This table shows the two-stage least squares estimates of the impact of an additional METCO student on suburban resident student college outcomes. The main endogenous variable is 20 times the average ratio of METCO to non-METCO students in an individual's grade cohort from grade 1. A one unit increase in the main endogenous variable represents an additional METCO student in a 20-student classroom. The model includes another endogenous variable: average class size excluding METCO students. This is estimated for 2001-2011 before administrative data on class size exists (see Equation 10). Linear and integer forms of predicted class size (see Equation 6 for estimation strategy) instrument for both the ratio of METCO to non-METCO students and non-METCO class size. Controls include binned total grade-level enrollment that proxy for the number of classrooms in a grade, school and year fixed effects, individual baseline covariates (including gender, race, free and reduced price lunch status, special education status, and English Language Learner status). Racial composition of college equals 0 if the student does not enroll in college and is for the first college the student enrolled in. The racial composition data comes from IPEDS annual data. Four-year college enrollment includes students who start in two-year colleges and transfer. Competitiveness rankings are determined by Barron's for the first four-year college the student enrolled in (if any). Panel B measures college enrollment within 18 months of the student's projected 4-year high school graduation. Panel D measures college persistence for on-time college enrollment (within 6 months of the student's projected 4-year high school graduation date).

Figure A.1: METCO Enrollment Over Time



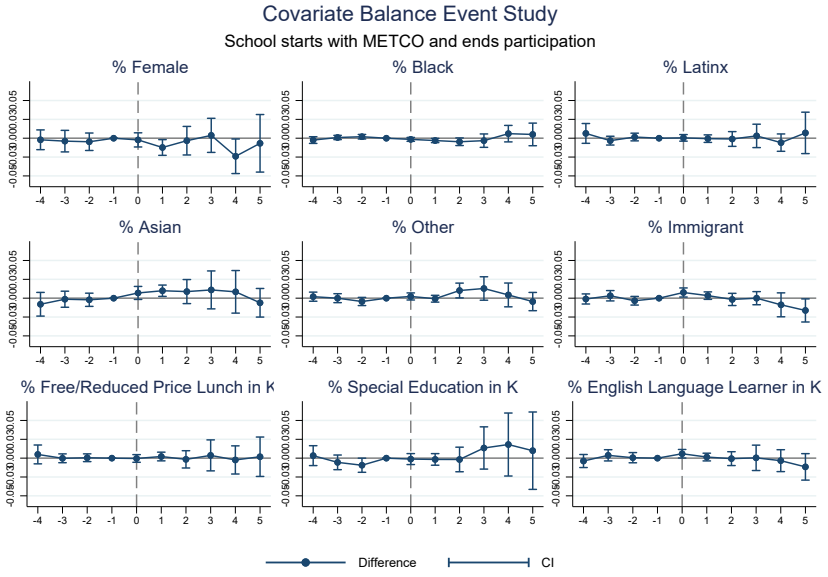
Notes: This figure plots the total METCO enrollment since the program started in 1966. Pre-2001 data comes from the Northeastern University Archives METCO Collection and was only available for Boston, not for Springfield. Not all years were available in the archives. The 2001 - present data come from the Massachusetts Department of Elementary and Secondary Education. All MA includes the Boston and Springfield programs.

Figure A.2: Event Study of Baseline Covariates



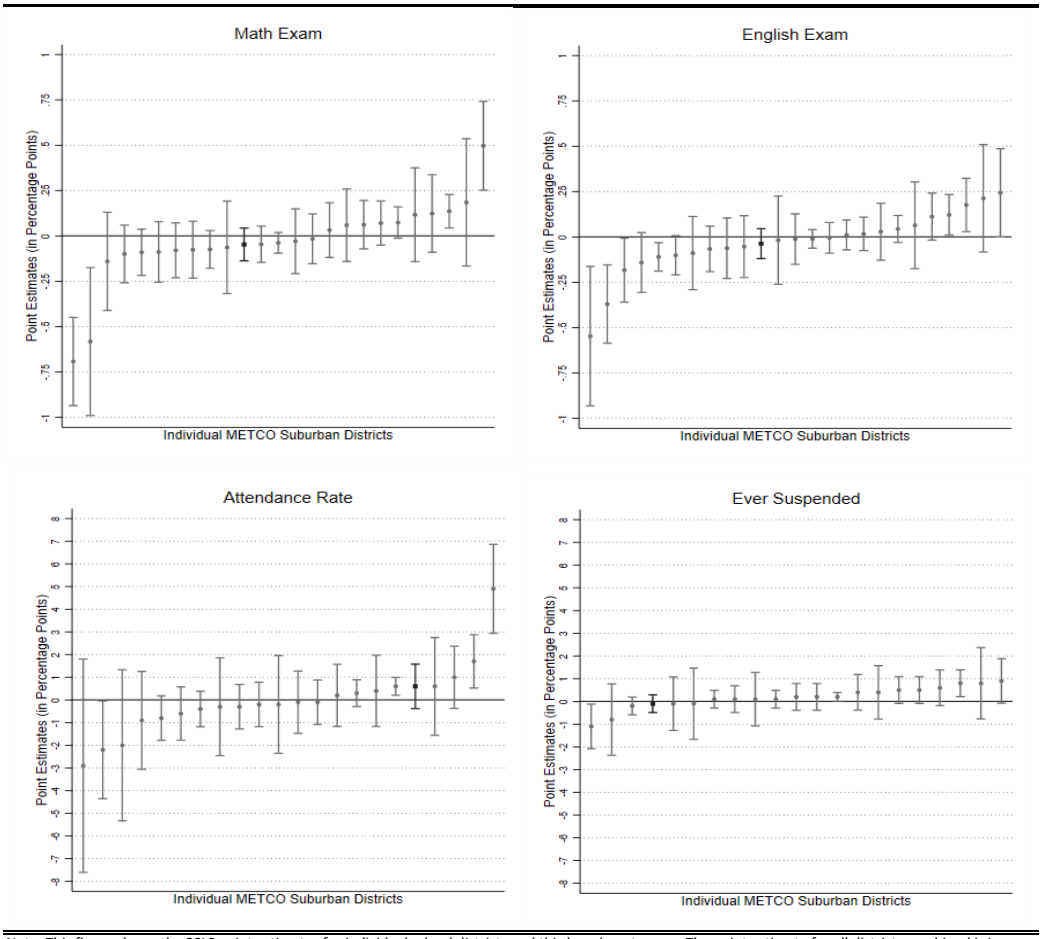
Notes: This figure plots the baseline covariates event study estimates from Equation 4 for schools that did not start with METCO and began enrolling METCO students.

Figure A.3: Event Study of Baseline Covariates



Notes: This figure plots the baseline covariate event study estimates from Equation 4 for schools that paused or ended their METCO enrollment.

Figure A.4: Individual School District Heterogeneity



Note: This figure shows the 2SLS point estimates for individual school districts and third grade outcomes. The point estimate for all districts combined is in black. See Table 5 for regression specifications.

Figure A.5: Map of Massachusetts METCO districts

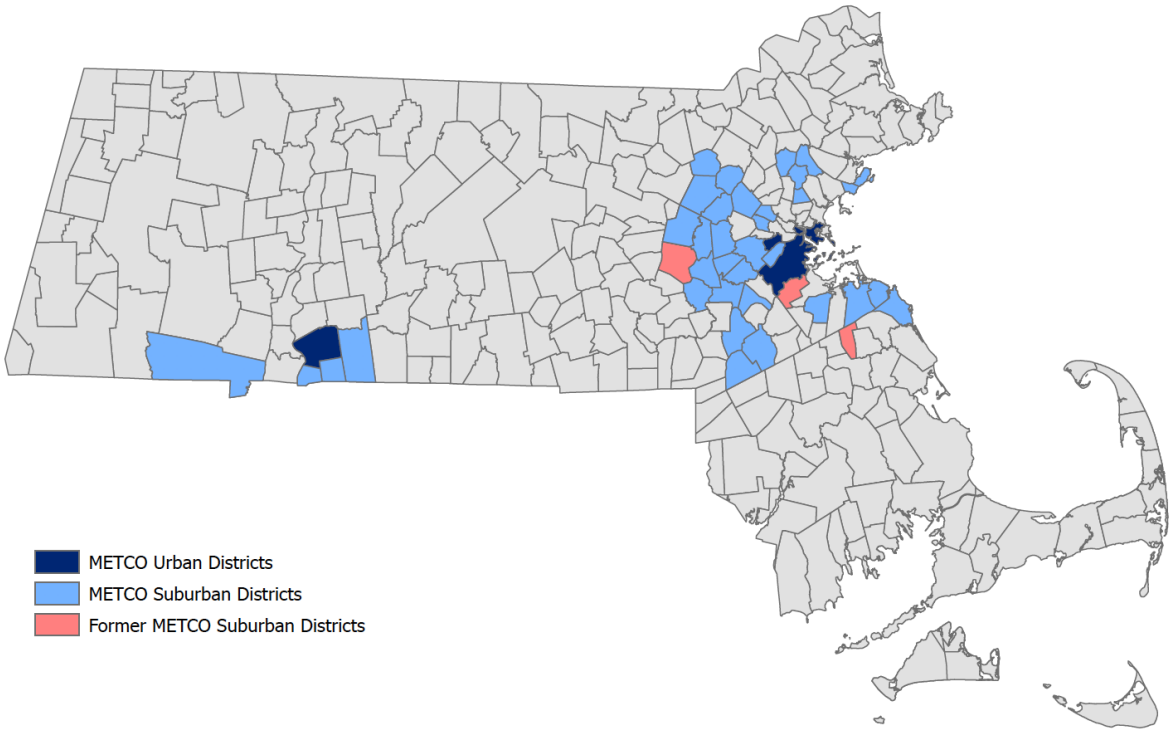


Table A.1: OLS Estimates of the Relationship between METCO and Black and Latinx Resident Student Outcomes

	Metro-Area			Receiving Districts				
	Non-METCO mean	Relationship between METCO and Outcome		Non-METCO mean	Relationship between METCO and Outcome			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Math	-0.523	0.451	0.042	-0.457	0.295	0.035	0.002	0.026
		(0.010)	(0.013)		(0.012)	(0.013)	(0.014)	(0.014)
N		659,821	659,821		81,705	81,705	71,481	67,608
English	-0.488	0.491	0.031	-0.385	0.303	0.024	-0.001	0.017
		(0.010)	(0.012)		(0.011)	(0.012)	(0.013)	(0.013)
N		652,537	652,537		80,661	80,661	70,500	67,671
Attendance Rate	0.814	0.025	0.005	0.811	0.015	0.005	0.002	0.004
		(0.001)	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
N		1,610,379	1,610,379		195,179	195,179	166,750	142,982
Suspended	0.088	-0.048	-0.009	0.046	-0.019	-0.001	0.001	0.000
		(0.001)	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)
N		1,551,174	1,551,174		189,003	189,003	166,750	142,982
Number of Suspensions	0.303	-0.193	-0.037	0.139	-0.075	-0.001	-0.001	-0.001
		(0.005)	(0.004)		(0.005)	(0.003)	(0.004)	(0.005)
N		1,551,174	1,551,174		189,003	189,003	166,750	142,982
Year& Grade FE, Region Control		X	X		X	X	X	X
District FE			X			X	X	X
Lagged School Traits							X	X
Individual Baseline Controls								X

Notes: This table reports the OLS estimates of the relationship between the proportion of METCO students in a schoolXgrade and the outcomes of Black and Latinx non-METCO students in that suburban district. The endogenous variable is the ratio of METCO to non-METCO students in a schoolXgrade multiplied by 20 so that a one unit increase in the endogenous variable can be interpreted as adding one METCO student to a 20-person class. All models control for whether the district is in the Springfield metropolitan area and include year and grade fixed effects. Lagged school traits include the prior year's average attendance rate, suspension rate, days suspended, and standardized Math and English test scores as well as the proportion of students receiving free or reduced price lunch, special education or English Language Learner services, and racial demographics. Individual baseline controls include the following information for the first year students appear in the data: gender, race and ethnicity, free and reduced price lunch, special education and level of services, English Language Learner and level of proficiency, and immigrant.

Table A.2: Descriptive Statistics of Analysis Samples

				Sporadic Sample (Narrow)		
	All Suburban Resident Students	DiD Sample				
Baseline Characterstics	(1)	METCO as Treatment (2)	No METCO as Treatment (3)	METCO as Treatment (4)	No METCO as Treatment (5)	2SLS Sample (6)
Female	0.49	0.49	0.49	0.49	0.48	0.49
Black	0.02	0.02	0.02	0.01	0.01	0.03
Latino/a	0.04	0.04	0.04	0.03	0.04	0.05
Asian	0.11	0.07	0.13	0.08	0.12	0.11
White	0.79	0.84	0.78	0.85	0.80	0.78
Subsidized Lunch	0.06	0.07	0.05	0.04	0.04	0.07
Special Education	0.07	0.08	0.09	0.08	0.08	0.07
Immigrant	0.02	0.01	0.03	0.01	0.02	0.02
English-Language Learner	0.06	0.03	0.06	0.03	0.05	0.06
English Exam	0.48	0.40	0.50	0.47	0.53	0.43
Math Exam	0.47	0.41	0.50	0.48	0.51	0.43
Unique Students	197,432	80,065	91,432	8,538	18,793	200,259
Unique Students w/ Test Scores	116,281	37,953	48,095	4,751	9,862	58,378
Years of Data	21	16	16	16	16	21
Unique Districts	35	30	31	18	22	35
Unique Districts Treated		25	25	18	22	
Unique Schools	152	117	102	31	59	154
Unique Schools Treated		78	83	31	59	
Unique School/Cohorts	2,842	1,165	1,287	126	272	
Unique School/Cohorts Treated		595	676	44	94	

Notes: This table displays baseline demographic characteristics for the original dataset and each of the analysis samples. Data is unique at the student-level. Math and English exam are the earliest exam scores (often third grade) and only include observations prior to the policy change for the difference and differences and sporadic samples and the non-treated observations for 2SLS.

Table A.3: Covariate Balance for Difference in Differences Analysis

	School does not start with METCO and Begins Participation			School starts with METCO and Ends Participation		
Baseline Characterstics	Treated Cohort Mean (1)	Difference between Treated Cohort and Untreated Cohort within Treated Schools (2)	Difference between Cohorts in Treated Schools and Untreated Schools (3)	Treated Cohort Mean (4)	Difference between Treated Cohort and Untreated Cohort within Treated Schools (5)	Difference between Cohorts in Treated Schools and Untreated Schools (6)
Female	0.484	0.007	-0.002	0.486	0.009	-0.005
	24,231	(0.006)	(0.004)	34,085	(0.005)	(0.002)
		42,664	1,412,621		52,018	1,903,735
Black	0.017	0.004	0.000	0.014	0.002	-0.001
	24,231	(0.001)	(0.001)	34,085	(0.001)	(0.001)
		42,664	1,412,621		52,018	1,903,735
Latino/a	0.041	0.002	0.001	0.037	0.001	0.000
	24,231	(0.002)	(0.001)	34,085	(0.002)	(0.001)
		42,664	1,412,621		52,018	1,903,735
Asian	0.114	0.001	0.005	0.111	-0.001	0.004
	24,231	(0.003)	(0.002)	34,085	(0.003)	(0.002)
		42,664	1,412,621		52,018	1,903,735
White	0.794	-0.004	-0.009	0.810	-0.001	-0.003
	24,231	(0.004)	(0.003)	34,085	(0.004)	(0.002)
		42,664	1,412,621		52,018	1,903,735
Subsidized Lunch	0.045	-0.001	0.000	0.039	0.000	0.002
	22,270	(0.002)	(0.002)	30,056	(0.002)	(0.001)
		38,453	1,269,866		46,462	1,695,728
Special Education	0.075	-0.001	0.001	0.085	-0.002	-0.002
	22,270	(0.003)	(0.002)	30,056	(0.003)	(0.001)
		38,453	1,269,866		46,462	1,695,728
English-Language Learner	0.053	0.000	0.002	0.048	0.002	0.002
	22,270	(0.003)	(0.002)	30,056	(0.002)	(0.001)
		38,453	1,269,866		46,462	1,695,728
Immigrant	0.020	0.001	0.001	0.019	0.000	0.001
	24,231	(0.002)	(0.001)	34,085	(0.001)	(0.001)
		42,664	1,412,621		52,018	1,903,735

Notes: This table shows descriptive statistics for the suburban resident students in cohorts that have METCO peers in Column (1). Column (2) shows the coefficients from regressions of observed characteristics on whether a student is in a treated cohort, controlling for school indicators in the sample of suburban resident students in schools that have METCO enrollment policy changes. Column (3) reports the coefficients from regressions of observed characteristics on whether a student is in a treated school versus a comparison school, controlling for each treatment and comparison group set.

Table A.4: Difference in Differences Robustness Checks

	School does not start with METCO and Begins Participation				School starts with METCO and Ends Participation			
	Main Specification	Minimum of Two Pre-Periods	Comparisons within Same District	Synthetic Control	Main Specification	Minimum of Two Pre-Periods	Comparisons within Same District	Synthetic Control
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Math	0.002	-0.028	0.018	0.034	-0.003	-0.044	0.006	-0.004
	(0.015)	(0.038)	(0.027)	(0.023)	(0.014)	(0.030)	(0.019)	(0.019)
Observations (Students)	990879	167170	25068	211551	1174017	132780	67098	282499
Treatment Groups	143	20	51	79	156	16	83	113
Control Schools	4653	603	88	778	5856	424	255	956
English	0.006	-0.024	0.014	-0.005	0.008	-0.052	0.002	0.001
	(0.015)	(0.037)	(0.032)	(0.022)	(0.013)	(0.024)	(0.017)	(0.015)
Observations (Students)	1104548	190761	27774	264539	1283026	138119	74791	351500
Treatment Groups	155	21	54	83	168	16	89	124
Control Schools	5101	633	93	976	6333	424	280	1262
Attendance Rate	0.000	0.002	-0.004	0.000	0.000	-0.002	0.000	0.000
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Observations (Students)	1126408	194535	28439	311833	1306911	140924	76273	429031
Treatment Groups	155	21	54	111	168	16	89	124
Control Schools	5103	633	93	1140	6337	424	280	1547
Suspended	0.000	0.002	-0.001	0.001	0.000	-0.002	0.002	0.000
	(0.001)	(0.002)	(0.001)	(0.000)	(0.001)	(0.002)	(0.001)	(0.001)
Observations (Students)	1126337	194525	28438	128346	1306906	140924	76273	130660
Treatment Groups	155	21	54	104	168	16	89	115
Control Schools	5103	633	93	568	6337	424	280	507
Number of Suspensions	-0.001	0.004	-0.002	0.001	0.002	0.003	0.004	0.001
	(0.001)	(0.003)	(0.002)	(0.001)	(0.001)	(0.005)	(0.002)	(0.002)
Observations (Students)	1126337	194525	28438	120299	1306906	140924	76273	139228
Treatment Groups	155	21	54	107	168	16	89	114
Control Schools	5103	633	93	516	6337	424	280	554

Table A.5: Broad Definition Sporadic Regression

	Effect of Being in a METCO Cohort					Effect of Being in No METCO Cohort				
	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Math			0.002	0.009	0.007			-0.010	-0.002	-0.017
			(0.019)	(0.022)	(0.021)			(0.015)	(0.013)	(0.016)
Observations (Students)			11,065	10,575	8,075			19,123	18,201	13,719
Treatment Groups			73	70	56			118	119	98
Control Schools			110	110	89			205	204	167
English			0.022	0.031	0.000			-0.007	-0.027	-0.026
			(0.020)	(0.020)	(0.022)			(0.014)	(0.014)	(0.016)
Observations (Students)			12,061	10,605	8,073			20,660	18,208	13,689
Treatment Groups			78	70	56			128	119	98
Control Schools			120	110	89			219	204	167
Attendance Rate	-0.002	0.002	-0.001	0.001	0.016	0.004	-0.001	0.000	0.003	-0.011
	(0.002)	(0.001)	(0.001)	(0.001)	(0.008)	(0.003)	(0.001)	(0.001)	(0.002)	(0.006)
Observations (Students)	13,238	12,688	12,258	11,654	9,275	23,990	22,915	21,013	20,247	17,004
Treatment Groups	78	78	78	77	63	130	130	128	128	117
Control Schools	120	120	120	118	101	223	223	219	219	201
Suspended	0.001	0.000	0.000	0.003	0.002	-0.001	0.000	0.000	-0.002	-0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
Observations (Students)	11,969	12,402	12,258	11,654	9,275	22,081	22,302	21,013	20,247	17,004
Treatment Groups	72	77	78	77	63	118	127	128	128	117
Control Schools	108	118	120	118	101	205	219	219	219	201
Number of Suspensions	0.004	-0.001	-0.001	0.004	0.002	-0.002	-0.001	0.002	-0.002	-0.003
	(0.002)	(0.001)	(0.002)	(0.003)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
Observations (Students)	11,969	12,402	12,258	11,654	9,275	22,081	22,302	21,013	20,247	17,004
Treatment Groups	72	77	78	77	63	118	127	128	128	117
Control Schools	108	118	120	118	101	205	219	219	219	201

Notes: This table shows the difference between cohorts with METCO relative to those without METCO after controlling for individual student baseline characteristics (gender, race, immigrant status, school fixed effects, year fixed effects. For Columns (1) through (5), the treatment is defined as one or two years of METCO enrollment that is preceded and followed by no METCO enrollment. The sample is restricted to these cohorts. In columns (6) through (10) the treatment and sample are defined as one or two years without METCO, preceeded and followed by cohorts with METCO.