# Supplement to "Still Worth the Trip? School Busing Effects in Boston and New York"

Joshua D. Angrist[*]    Guthrie Gray-Lobe [†]    Clemence Idoux [‡]    Parag Pathak[§]

July 2022

## Abstract

This file contains four online appendices. Appendix B contains additional notes for Figure 2. Appendix C details the construction of measures of school proximity. Appendices D and E describe the Boston and New York data used in the paper.

---

[*]MIT Department of Economics and NBER. Email: angrist@mit.edu

[†]University of Chicago Department of Economics. Email: graylobe@uchicago.edu

[‡]MIT Department of Economics. Email: cidoux@mit.edu

[§]MIT Department of Economics and NBER. Email: ppathak@mit.edu

# B    Additional Notes to Figure 2

We downloaded the most recent version posted of the Public Elementary/Secondary School Universe Survey Data file from the National Center for Education Statistics (NCES) website on May 6, 2021. We adjusted school-level enrollment by race as described in the errata accompanying the data documentation. This yields year-specific files containing enrollment characteristics for every public school in the U.S. and territories.

For school years 1988-89 through 2000-01, enrollment by race was retrieved using the race-specific student count variable. For school years 2001-02 through 2018-19, enrollment by race is calculated by collapsing student count by race, grade, and gender to enrollment by race. For school years 2016-17 through 2018-19 the variable changes name and is labelled as Category Set A - By Race/Ethnicity; Sex; Grade.

Online schools are by eliminating those with a virtual status of *"Yes"*, *"Full Virtual"*, *"Supplemental Virtual"*, and *"Virtual with face-to-face options"*. Schools marked as closed are also omitted. Student counts given for a grade level of *"Adult Education"* are also omitted.

Total school enrollment is defined as the sum of all reported enrollment by race. Schools not reporting enrollment by race are therefore omitted. In 2009, NCES changed race reporting to allow schools to specify *"Two or more races"*, *"Native Hawaiian or other Pacific Islander"*, and *"Not specified"*. This contributes to changes in peer shares by race over time.

Boston schools are those where the local education agency (LEA) identifier is 2502790. New York City schools are identified by a single LEA prior to the 2005-06 school year and by community district LEAs in later years. New York LEAs are 3620580 (the aggregated district), 3600135 (D75 special needs), 3600075 (alternative schools), or one of the 32 community district LEAs (3600076, 3600077, 3600078, 3600079, 3600081, 3600083, 3600084, 3600085,3600086, 3600087, 3600088, 3600090, 3600091, 3600119, 3600092, 3600094, 3600095, 3600096, 3600120, 3600151, 3600152, 3600153, 3600121, 3600098, 3600122, 3600099, 3600123, 3600100, 3600101, 3600102, 3600103, 3600097). This definition of New York schools follows NCES documentation.

Racial isolation measures district-level exposure to peer share Black, Hispanic, from the perspective of a given race, coded by $r_p \in \{$Black, Hispanic, Black or Hispanic$\}$. Racial groups defining exposure are coded by $r_e$ chosen from the same set as $r_p$. The number of $r_p$ and $r_e$ students in the district is denoted $N_p$ and $N_e$. Districts are made up of

- a set of schools $s \in S$ enrolling $n_s$ students

- School- and race-specific enrollment counts $n_{sp}$ and $n_{se}$.

Exposure to racial group $r_e$ for students in racial group $r_p$ is computed as:

$$e_{pe} = \sum_{s \in S} \left( \left( \frac{n_{se}}{n_s} \right) \cdot \left( \frac{n_{sp}}{N_p} \right) \right).$$

Minority isolation rates are computed as follows. Let $t \in \{80, 90\}$ denote the cutoff defining minority isolation and let $n_{sm}$ denote the number of minority (Black or Hispanic) students enrolled at school $s$. Indicate isolated schools by $D_s = 1[\frac{n_{sm}}{n_s} > t]$. Minority isolation of race group $r_p$ in a district is then

$$I_{pt} = \sum_{s \in S} D_s \frac{n_{sp}}{N_p}.$$

## C   Measuring School Proximity

Our econometric framework uses two measures of school proximity, one based on distance and one based on travel time. For Boston students, proximity is measured by a dummy indicating school attendance inside applicants' Garrity walk zone and by travel time in 20-minute units. For NYC students, proximity is measured by a dummy indicating school attendance inside applicants' residential district and by travel time in 20-minute units.

### Garrity Walk Zone Offer/Enrollment Status

For Boston, non-neighborhood offers and enrollment are defined by whether schools offered or attended are outside applicants' and students' residential walk zone. A Boston student is said to reside within a school's walk zone if a one-mile-radius circle around the school intersects the boundary defining the student's residential geocode. BPS codes walk-zone status only for applicants (this is a priority in the match). We use this data to identify the set of residential geocodes in each school's walk zone, thereby imputing walk zone enrollment status for all students in the cohorts analyzed here. Charter and exam schools are coded as out-of-walk zone for all students.

### New York District Offer/Enrollment Status

For New York, non-neighborhood offers and enrollment are defined by whether schools offered or attended are outside applicants' and students' residential district. New York City schools are allocated to one of 32 community districts that partition all addresses in the city. School districts are identified by the first two digits of the school DBN. Charter and Specialized high schools are coded as being out-of-district for all applicants.[1]

### Travel Time

Travel time measures the time it takes to travel from residential GPS coordinates to school GPS coordinates. Boston school coordinates are inferred from addresses. New York school coordinates for all non-charter schools appear in NYCDOE High School Directory files.[2] New York charter school coordinates are inferred from school addresses. The student data sets we work with omit

---

[1]Similarly, a small set of alternative schools are coded out-of-district for all applicants. See `https://infohub.nyced.org/in-our-schools/programs/district-79` for list of alternative schools.

[2]These were obtained online from `https://opendata.cityofnewyork.us`.

home address. Residential coordinates are approximated by the coordinates of the centroid of residential geocodes in Boston and by the centroid of residential census tracts in New York.

Travel times are obtained by applying the HERE Public Transit API to pairs of (home,school) coordinates.[3] Specifically, we compute minutes of travel with an arrival time of 8:00 AM on Monday, January 31st, 2022. The resulting times are winsorized at the 1st and 99th percentiles of the relevant distribution and divided by 20.

# D   Boston Data Appendix

Lists of middle school and high school applicants, applicant rank order lists, and assignments are constructed using annual records from the Boston Public Schools (BPS) school assignment system. Information on student demographic characteristics and schools attended comes from the Student Information Management System (SIMS), a centralized database that covers all public school students in Massachusetts. Achievement test scores are from the Massachusetts Comprehensive Assessment System (MCAS). College attendance information comes from the National Student Clearinghouse (NSC). Data on school characteristics were obtained from publicly available Massachusetts Department of Elementary and Secondary Education (DESE) online records. SIMS and NSC data were provided by DESE. MCAS data from 2000-2001 are from Boston and MCAS data from 2002-2017 are from DESE. Other Boston data sources are detailed below.

### BPS Assignment Data

We received assignment data from BPS for all applicants to 6th and 9th grade seats in traditional and some pilot schools between 1997-2020. These files exclude applicants who applied only to selective enrollment (exam) schools and/or charter schools. The data include preferences as reflected in rank order lists, priorities at each program listed, lottery tiebreaker values, and the program to which the applicant was assigned. In the assignment data, applicants rank programs rather than schools; many schools host multiple programs, including bilingual and special education programs. Assignment data include applicant geocode, defined as one of 867 Garrity-era geographic areas that determine walk-zone priority in the contemporary Boston match.

The analysis sample covers applicants for 6th and 9th grade seats in the school years beginning 2002-2013. DESE data on MCAS outcomes begins in 2002. Fall 2014 marked the introduction of a new assignment scheme known as the Home-Based Assignment Plan (HBAP). The HBAP phased out walk zone priorities and limited the set of schools applicants may rank, favoring neighborhood schools while still including at least some with good outcomes (Shi, 2015; Pathak and Shi, 2021).

Depending on the year, the BPS assignment system has three or four rounds. Most students (87-95% of 6th graders and 83-93% of 9th graders in years 2002-2013) participate in the first round. Our analysis covers first-round participants only. Application counts by year and grade appear in columns 1 and 3 of Table B1. Preferences, priorities, and capacities for each program (the last

---

[3]See https://developer.here.com/documentation/public-transit/dev_guide/index.html.

of which is computed by summing the number of offers) are used to replicate the match. The steps from priorities to rankings over students are taken from Pathak and Shi (2014). Our match replication rates mostly exceed 95%, a fact documented in columns 2 and 4 of Table B1.

Application data are supplemented by BPS enrollment records. This file provides geocode data for application years in which geocodes are missing from the assignment files (2006-07 to 2010-11). Application and enrollment files are linked by a BPS-defined applicant identifier.

## SIMS Data

Our analysis uses SIMS files for the school years beginning fall 2002-13. These files include information on student demographics, behavioral outcomes, and schools attended for all students enrolled in Massachusetts public schools each year. SIMS contain a state student identifier known as the SASID, used to match SIMS records to MCAS and NSC files.

SIMS data are used to code school enrollment, special education status, subsidized lunch status, and limited English proficiency. Students who attended multiple schools in the same school year are recorded as attending the school they attended longest. Students are classified as qualifying for special education services or eligible for a free or reduced-price lunch in any record within a school-year-grade retain that designation for the entire year-grade. SIMS data are also used to code behavioral outcomes, including ever being suspended, number of suspensions, ever being truant, number of days truant, ever attending school operated by the Massachusetts Department of Youth Services, and number of days absent. SIMS omits data on suspensions after 2012-13. Suspension data for the 2013-14 cohort come from School Safety and Discipline Reporting (SSDR) files provided by DESE.

## MCAS Data

The Massachusetts Comprehensive Assessment System (MCAS) is the state accountability exam for Massachusetts, taken in grades 3-10. Our analysis uses MCAS Math and English Language Arts (ELA) data from the school years beginning 2000-2017. MCAS scores are linked with student SIMS records using SASIDs. Our analysis uses MCAS ELA and Math test scores from the 4th, 6th, 7th, 8th, and 10th grade exams. Grade 4 Math and ELA scores are used as baseline controls for applicants to grade 6. Grade 8 Math scores and grade 7 and 8 ELA scores are used as baseline controls for applicants to grade 9. As as baseline, grade 7 ELA scores from years 2000-01 through 2003-04 are used for applicants enrolled in 9th grade in school years 2002-03 through 2005-06 and grade 8 ELA scores from years 2005-06 through 2012-13 are used for applicants enrolled in 9th grade in school years 2006-07 through 2013-14. Grade 6 Math scores and grade 7 ELA scores are the MCAS outcomes for applicants to grade 6. Grade 10 Math and ELA scores are the MCAS outcomes for applicants to grade 9. Raw scores are standardized to have mean zero and standard deviation one within a subject-grade-year among enrolled students in our sample who are tested in a given year.

4

### NSC Data

Data on college enrollment comes from the National Student Clearinghouse (NSC) database, which captures enrollment for 96% of U.S. college student Hindley and Eaton (2018). The Massachusetts DESE regularly requests an NSC extract with college-going information for Massachusetts high school graduates and dropouts (requests are mostly made annually for graduates and every two years for dropouts and other non-graduates). The resulting matched data set includes each student-school enrollment record found in the NSC database. Student enrollment records are associated with SASIDs, thereby linking NSC data to SIMS data. We use NSC records from the 2006-07 through 2017-18 school years, covering potential college enrollment for 9th graders in our sample. Students are coded as having enrolled in college if they appear as enrolled in NSC data within 6 months of their projected high school graduation. NSC college enrollment data distinguish between four-year and two-year schools.

### Boston School Characteristics

Data on school characteristics come from publicly available records posted online.[4] These records provide school-specific information on student-teacher ratios, the percent of teachers licensed in their teaching assignment, and the percent of core academic classes taught by highly-qualified teachers. Information on these three outcomes is available for all cohorts in our sample. The student-teacher ratio is defined as the total number of students divided by the number of full-time equivalent (FTE) teachers across all grade levels at the school attended. The number of FTE teachers counts all teaching professionals, including teachers who work with special populations such as high needs students or English learners in small groups. The percent of teachers licensed in their teaching assignment refers to the percent of teachers who are licensed with provisional, initial, or professional licensure to teach in the subject(s) in which they are posted. Finally, the percent of core academic classes taught by highly-qualified teachers gives the percent of core subjects (English language arts, mathematics, and science, among others) taught by teachers holding a Massachusetts teaching license and demonstrating subject matter competence in the areas they teach.

## E  New York Data Appendix

Lists of high school applicants, applicant rank order lists, and assignments are constructed using annual records from the New York City Department of Education (NYCDOE) school assignment system. Information on student demographic characteristics and schools attended comes from the NYCDOE's Office of School Performance and Accountability (OSPA). Achievement test scores are taken from New York State Assessments and the SAT. College attendance information comes from the National Student Clearinghouse (NSC). Data on school characteristics were obtained from publicly available school report cards posted online. Geographic information on students

---

[4]See https://profiles.doe.mass.edu/statereport/teacherdata.aspx.

and schools comes from an administrative data set known as the Zoned District Borough Number (DBN) file. All non-public data were provided by NYCDOE. These files include a unique student identification number for record linkage. New York data sources are detailed below.

## NYCDOE Assignment Data

Data on NYC high school applications were provided by the NYCDOE enrollment office. We received assignment data for all applicants seeking 9th grade seats at traditional city high schools for the school years beginning 2012-16. These files exclude applicants who applied only to selective enrollment (exam) schools and/or charter schools. Application records include students' rank-order lists of academic programs submitted in each round of the application process, priority and rank at each program listed, lottery tiebreaker values, and the program to which the applicant was assigned.

We focus on 9th grade applicants seeking seats in school years beginning fall 2012-2016, and who submitted preferences in the main round of the high school match. The NYC high school match includes three rounds. The main round uses a deferred acceptance (DA) algorithm. Between 85-89% of applicants in years 2012-2016 were assigned in the main round. Applicants unassigned in this round typically apply in a subsequent round.[5] Students who remain unassigned in later rounds are assigned on a case-by-case basis in a final administrative process. Application counts by year appear in Column 1 of Table B2. Preferences, priorities, and capacities for each program (computed by summing the number of observed offers) are used to replicate the match. As can be seen in column 2 of Table B2, we replicate an average of 96.0% or more of main round assignments.

## OSPA and Address Data

We received registration and enrollment files for the school years starting fall 2012-16 from NYC-DOE's OSPA office. These data include students' grade and enrolled-school DBN, as well as student demographic variables. OSPA data for a given year include a record for every student enrolled at a New York City public school any time during the year. Our data come from end-of-year files recorded in June. This file supplies information on school enrollment, attendance, special education status, subsidized lunch status, and limited English proficiency.

DBN files contain a residential address for most New York City students. We search DBN files for 2010-2018 when address information is missing in cohort-specific files.

## New York State Assessment Data

Lagged test score controls come from New York State Assessments, a set of state accountability tests covering grades 3-8. Our lagged score controls come from 6th grade tests in math and English Language Arts (ELA) scores, taken in spring 2010 through spring 2014. Baseline scores are nor-

---

[5]Students in the first round are those who also apply to the city's specialized high schools. If a student does not receive a specialized high school offer, they are notified of their offer in the main round. See Abdulkadiroğlu, Pathak and Roth (2005) for details.

malized to have mean zero and standard deviation one within a subject-year among all 9th grade public school students in the city.

## SAT Data

The NYCDOE supplied data on SAT scores covering the period in which our applicant cohorts were tested. NYCDOE obtains SAT data for New York City students from the College Board, which owns and administers the SAT. College board data are matched to DOE files using name and date-of-birth.

Many students take the SAT repeatedly. For those that do, we use the score from the first test. Most of these were taken in 11th grade. SAT score scales were redesigned in 2016. We therefore scale pre-redesign scores using the detailed in `https://collegereadiness.collegeboard.org/educators/higher-ed/scoring/concordance` for the conversion scale. Raw scores are standardized to have mean zero and standard deviation one within a subject-year among enrolled students in our cohorts who are tested in a given year.

## NSC Data

As for Boston, New York college outcomes come from the NSC. New York NSC data come to us via NYCDOE, which has NSC data for cohorts graduating through 2019. Our college outcomes therefore omit the 2016 applicant cohort. College graduation and degree type are coded as for Boston.

## New York School Characteristics

Data on school characteristics come from publicly available files released online by the New York State Education Department.[6] We use three outcomes from these data files: the student-teacher ratio, the percent of teachers licensed in their teaching assignment, and the percent of core academic classes taught by highly-qualified teachers. The student-teacher ratio is the total number of students divided by the number of full-time equivalent (FTE) teachers across all grade levels at the school attended. The number of FTE teachers counts all teaching professionals, including teachers who work with special populations, such as high needs students and English learners, in small groups. This data is available for school years 2012-2016. The percent of teachers licensed in their teaching assignment refers to the percent of teachers who are licensed to teach in the subject(s) in which they are posted. This data is also available for school years 2012-2016. Finally, the percent of core academic classes taught by highly-qualified teachers gives the percent of core subjects that are taught by teachers holding a New York State teaching license and demonstrating subject matter competence in the areas they teach. This data is available for 2012-2015 only.

School addresses are obtained from the Zoned DBN file.

---

[6]See `https://data.nysed.gov/downloads.php`.

# References

**Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth.** 2005. "The New York City High School Match." *American Economic Review, Papers and Proceedings*, 95(2): 364–367.

**Hindley, B., and V. Eaton.** 2018. "Staying the Course: Six-Year College Enrollment and Completion of BPS Class of 2011 Graduates." The Boston Private Industry Council.

**Pathak, Parag A., and Peng Shi.** 2014. "Demand Modeling, Forecasting, and Counterfactuals, Part I." NBER Working Paper No. 19859.

**Pathak, Parag, and Peng Shi.** 2021. "How Well Do Structural Demand Models Work? Counterfactual Predictions in School Choice." *Journal of Econometrics*, 222(1A).

**Shi, Peng.** 2015. "Guiding School-Choice Reform through Novel Applications of Operations Research." *Interfaces*, 45(2).

Table B1: Boston Match Replication Rates

| | Middle School | | High School | |
|---|---|---|---|---|
| | Number of 1st round applicants | Replication rate | Number of 1st round applicants | Replication rate |
| | (1) | (2) | (3) | (4) |
| 2002 | 4,052 | 95.5% | 3,098 | 94.5% |
| 2003 | 4,239 | 96.1% | 3,143 | 95.2% |
| 2004 | 3,868 | 94.4% | 3,416 | 95.3% |
| 2005 | 3,596 | 97.4% | 3,323 | 94.3% |
| 2006 | 2,884 | 97.5% | 3,062 | 99.8% |
| 2007 | 2,466 | 96.4% | 2,714 | 100.0% |
| 2008 | 2,149 | 97.4% | 2,609 | 100.0% |
| 2009 | 2,352 | 96.1% | 2,738 | 99.7% |
| 2010 | 2,320 | 95.5% | 2,691 | 96.6% |
| 2011 | 2,074 | 97.5% | 2,495 | 93.8% |
| 2012 | 2,073 | 96.6% | 2,727 | 93.3% |
| 2013 | 1,748 | 95.5% | 2,810 | 96.3% |

Table B2: NYC Match Replication Rates

| | High School | |
|---|---|---|
| | Number of 1st round applicants | Replication rate |
| | (1) | (2) |
| 2012 | 67,778 | 96.8% |
| 2013 | 66,342 | 96.0% |
| 2014 | 68,077 | 96.6% |
| 2015 | 64,911 | 98.2% |
| 2016 | 64,698 | 97.8% |