Do Boston's Schools Look Like Their Neighborhoods?

Samantha Batel, Emma Davies, and Sam Horan

December 4, 2018

API 201Z Final Exercise

Executive Summary

Boston's public schools have been criticized for not representing the communities they serve, with stakeholders alleging that the student population is composed disproportionately of students of color. To address this critique, this analysis investigates whether or not (1) Boston schools reflect their neighborhoods' diversity and (2) school or neighborhood demographics predict student achievement after controlling for other basic school characteristics, such as student-teacher ratio. Using 2015 data from the U.S. Census Bureau's American Community Survey and the Boston Public Schools, we find that (1) Boston schools do not look like their neighborhoods, containing significantly more students of color than one would expect; (2) school demographics have some significance as predictors of student achievement, whereas neighborhood-demographic factors generally do not; and (3) school type—charter vs. traditional public school—is a significant explanatory variable of student achievement.

Introduction

In Boston, a majority-minority city, neighborhoods are racially segregated.[1] Within this context, the city uses a neighborhood schools model to assign students in grades K-8 to schools based on where they live.[2] In spite of overall segregation in the city, then, these schools should presumably reflect their neighborhoods' racial and ethnic demographics.

Our analysis first explores whether Boston's schools do in fact look like their surroundings, comparing the racial and ethnic composition of schools to those of their neighborhoods. Next, given the historical association between race/ethnicity and performance—in part attributable to a resource gap—this analysis investigates whether school or neighborhood demographics predicts student performance. The analysis also considers additional factors that may affect student achievement, including the proportion of economically disadvantaged students in a school; student-teacher ratio; and type of school (traditional, exam, or charter).

Description of Data

We are primarily interested in the racial/ethnic demographics of 105 Boston schools and the 17 Boston neighborhoods that contain them. Our study uses 2015 data from the U.S. Census Bureau's American Community Survey[3] and the Massachusetts Department of Elementary and Secondary Education (DESE). The census data provide neighborhood demographics, while DESE provides school-level information on school type, demographics, student-teacher ratio, student economic disadvantage,[4] and achievement. Student achievement is measured by

---

[1] Katie Johnston, "Around Massachusetts, racial divides persist," *Boston Globe*, April 17, 2017, available at https://www.bostonglobe.com/business/2017/04/17/around-massachusetts-racial-divides-persist/HqQrm3TcH1od1j2qQ2F44J/story.html.

[2] Some students in our data attend schools that also serve high schoolers, but all students analyzed are in grades K-8.

[3] The census data were aggregated at the neighborhood level by Analyze Boston, the City of Boston's open data hub. Data available at http://www.bostonplans.org/getattachment/8aefd3d3-81ae-4a76-bb9c-9e2d6a3fa8c3.

[4] Boston Public Schools defines "economically disadvantaged" as being enrolled in "the Supplemental Nutrition Assistance Program (SNAP); the Transitional Assistance for Families with Dependent Children (TAFDC); the Department of Children and Families' (DCF) foster care program; and MassHealth (Medicaid)." Massachusetts DESE, "Information Services – Data Collection," available at http://www.doe.mass.edu/infoservices/data/ed.html.

standardized student scores from the Massachusetts Comprehensive Assessment System (MCAS). The demographics of the individual test takers are not available.

Figures 1 and 3 below show the mean Math and English MCAS scores for schools plotted against the proportion of white students in the school, while Figures 2 and 4 show the mean test scores for schools plotted against the ratio of white students in the school to white residents in the school's neighborhood. The scatterplots appear to show a relationship between demographics and mean test scores, particularly for traditional schools. Table 1 summarizes our variables of interest for the five largest neighborhoods and in aggregate across all 17 neighborhoods. The school demographic percentages are a weighted average for the neighborhood based on the size of the school population. Exam scores are weighted averages based on the number of test-takers.

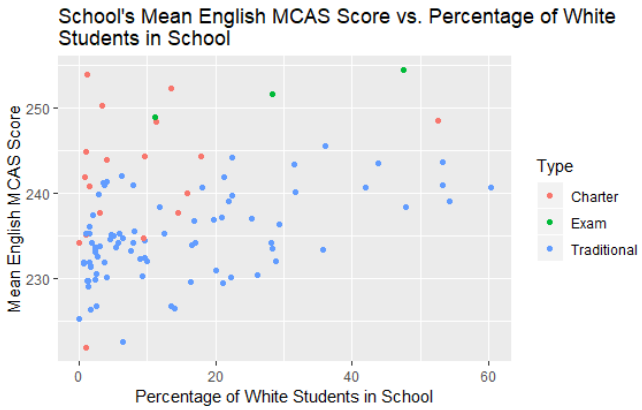

Figure 1

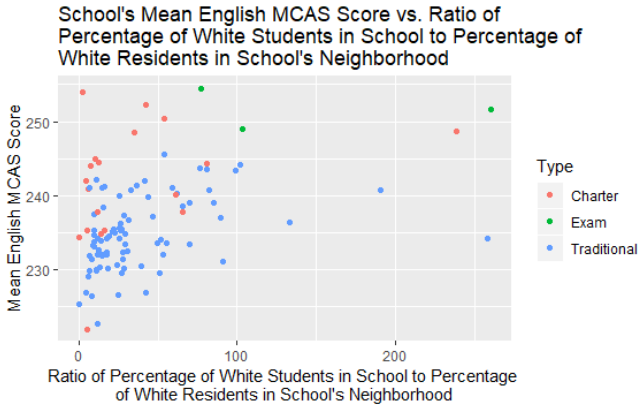

Figure 2

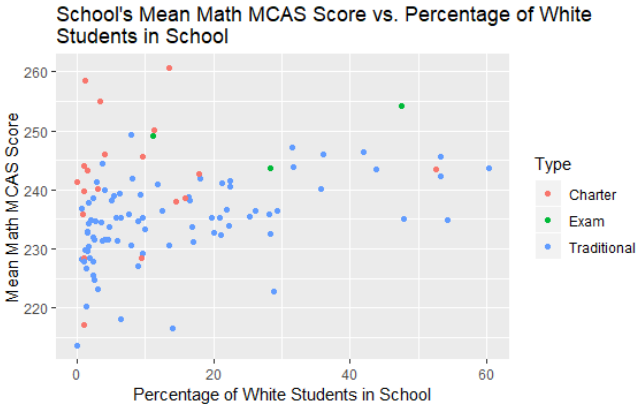

Figure 3

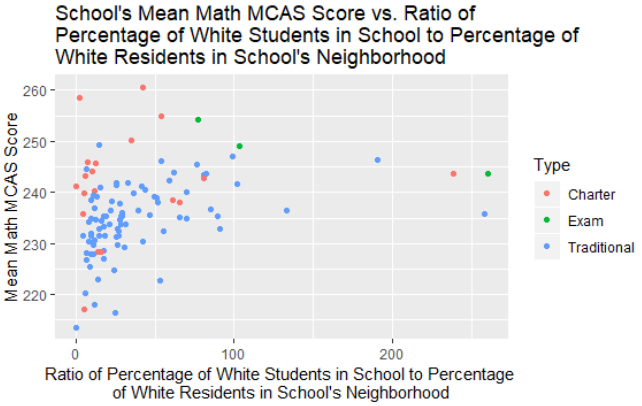

Figure 4

Table 1

	5 Largest Neighborhoods					All Neighborhoods
	Dorchester	Roxbury	Brighton	East Boston	Jamaica Plain	
NEIGHBORHOOD DEMOGRAPHICS						
Population	124,489	51,252	47,768	44,989	39,240	650,281
Race/Ethnicity						
White	22.1%	10.9%	66.9%	32.0%	54.3%	45.5%
Black	44.1%	53.0%	4.1%	2.4%	11.6%	22.7%

Hispanic	16.3%	29.0%	10.4%	57.9%	25.4%	18.8%
Asian/Pacific Islander	10.6%	3.2%	15.4%	3.6%	5.5%	9.3%
Other	6.9%	3.9%	3.2%	4.1%	3.2%	3.7%
SCHOOL DEMOGRAPHICS						
Race/Ethnicity						
White	10.2%	9.8%	18.6%	12.4%	19.1%	14.4%
Black	45.5%	37.8%	29.9%	6.1%	21.5%	34.8%
Hispanic	31.9%	40.8%	37.9%	77.9%	52.9%	39.3%
Asian	9.0%	8.8%	10.4%	1.8%	1.9%	8.3%
Other	3.4%	2.8%	3.2%	1.8%	4.6%	3.2%
Student teacher ratio (avg.)	13.4	15.0	12.6	13.9	14.3	13.9
Pct. Economically disadvantaged Students	53.5%	49.0%	47.8%	47.5%	46.5%	48.5%
Average Math Scores	233.8	237.5	233.7	240.6	235.5	237.3
Average Reading Scores	236.3	240.0	235.2	238.5	237.6	238.1
Number of Charter Schools	6	2	1	2	0	19
Number of Exam Schools	0	2	0	0	0	3
Number of Traditional Schools	19	10	3	8	5	83
Total Number of Schools	25	14	4	10	5	105

Description of Statistical Methods

We perform a two-stage statistical analyses, first comparing the racial demographics of a neighborhood's schools to the demographics of the neighborhood overall, then using regression techniques to investigate the relationship between various neighborhood and school characteristics and that school's MCAS results.

To compare school demographics to neighborhood demographics, we aggregated school data by neighborhood to calculate the number of white students in that neighborhood. We then calculated the expected number of white students based on neighborhood demographics and student population, and compared this expected number to the observed total using a chi-squared test. Finally, we conducted individual hypothesis tests for each neighborhood to compare whether the proportion of white students in individual neighborhoods' schools differed from the proportion of white residents in the neighborhood in a statistically significant way.

Next, we analyzed which factors predict a school's mean MCAS scores. We conducted four ordinary least-squares regressions, two each using schools' mean math and English MCAS scores as response variables. For each MCAS test type, we performed one regression using only school characteristics—including racial representation, economic disadvantage, student-teacher ratio, and school type (traditional public, charter, or exam)—as explanatory variables. We followed this with a second regression that included neighborhood demographic data and the ratio of the proportion of white students in the school to the proportion of white residents in the neighborhood as explanatory variables to examine whether the relationship between school and neighborhood demographics had any predictive significance.

The smaller regressions produced models of the form:

$$\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1}X_{Exam\ School} + \widehat{\beta_2}X_{Traditional\ School} + \widehat{\beta_3}X_{Prop.White\ Students} + \widehat{\beta_4}X_{Prop.Asian\ Students} + \widehat{\beta_5}X_{Prop.Hisp.Students} + \widehat{\beta_6}X_{ST\ Ratio} + \widehat{\beta_7}X_{Prop.Ec.Dis.}$$

where $\hat{Y}$ is the estimated mean MCAS score, the $\widehat{\beta_i}$ are estimated coefficients, and the X_i are the various explanatory variables. The larger regressions produced models of the form:

$$\hat{Y} = \widehat{\beta_0} + \widehat{\beta_1}X_{Exam\ School} + \widehat{\beta_2}X_{Traditional\ School} + \widehat{\beta_3}X_{Prop.White\ Students} + \widehat{\beta_4}X_{Prop.Asian\ Students} + \widehat{\beta_5}X_{Prop.Hisp.Students} + \widehat{\beta_6}X_{ST\ Ratio} + \widehat{\beta_7}X_{Prop.Ec.Dis.} + \widehat{\beta_8}X_{Prop.White\ Residents} + \widehat{\beta_9}X_{Prop.Asian\ Residents} + \widehat{\beta_{10}}X_{Prop.Hisp.Residents} + \widehat{\beta_{11}}X_{Ratio\ of\ White\ Students\ to\ White\ Residents}$$

Results

The first part of our analysis finds a statistically significant difference between the racial makeup of Boston's schools and the racial makeup of its neighborhoods. Specifically, the schools in all of Boston's neighborhoods contain a substantially lower proportion of white students than the proportion of white residents in the surrounding neighborhood.[5] The discrepancy is visually stark, as shown in Figure 5, at right.

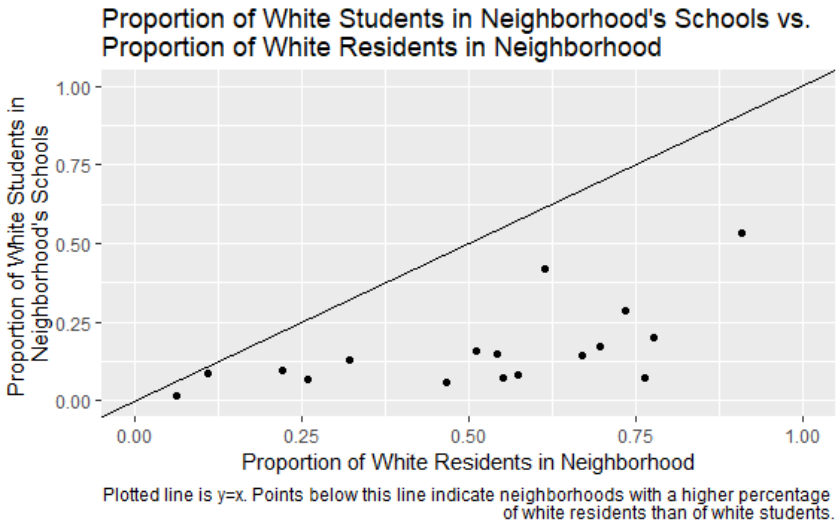

Figure 5

The second part of the analysis indicates that several of the explanatory variables we considered were linked to changes in mean MCAS scores in statistically significant ways. Taken together, variation in these explanatory variables accounted for roughly 40% to 50% of the variation in mean test scores, depending on which model one considers.

The type of school—in particular, charter vs. traditional—was a significant explanatory variable in all four regression models. Though the estimated coefficient varied between models, the results of the regression indicate that a school's status as a charter school, as opposed to a traditional public school, corresponds on average to an increase of between 7.22 and 9.85 (depending on which regression model one looks at) in the school's mean MCAS score, holding the other explanatory variables considered here constant.

The school's demographics also had some significance as predictors of MCAS performance. The proportion of Asian students in the school, in particular, was associated with positive changes in the schools' MCAS performance. Holding other factors constant, the four regression models link an increase of one percent in the proportion of Asian students to an increase of between 0.15 and 0.32 points in the school's MCAS results, on average. This coefficient was significant at the .01

[5] This neighborhood-by-neighborhood discrepancy is true of traditional public schools. Charter and exam schools were excluded from this portion of the analysis because their students are not assigned on a geographic basis.

level in all four tests. Other demographic information was significant in some models but not in others: the proportion of white students, the proportion of Hispanic students, and the proportion of economically disadvantaged students were each significant (though only at the relatively weak .10 level) in one of the four models.[6] Taken together, these findings suggest that a school's demographic characteristics may be linked to changes in the school's MCAS performance, even when controlling for other factors, such as school type.

Perhaps unsurprisingly—in light of the differences between schools and their neighborhoods discussed above—neighborhood-level demographic factors were generally not significant predictors of schools' MCAS results. The proportion of Asian residents in a neighborhood was an exception, though it was significant only at the relatively weak .10 level. Interestingly, this relationship pointed in the opposite direction than for Asian students, as the models linked an increase of one percent in the neighborhood's proportion of Asian residents to an average decrease of .30 and .24 points in a school's average math and English MCAS scores, respectively, holding other explanatory factors constant. Because of the relatively low significance level of this relationship, and the fact that no other neighborhood demographic factors were found to be significant predictors of MCAS results, it appears that neighborhood-level factors correspond less closely to school MCAS results.

Conclusion

This analysis shows a significant demographic mismatch between Boston's schools and its neighborhoods and suggests that this mismatch may have implications for the city's education policy. In each neighborhood that is home to a K-8 school, there are proportionally fewer white students compared with the proportion of white residents. Although the data and techniques used here cannot independently provide a causal explanation for this discrepancy, these findings support the claim that Boston schools are even less diverse than the city's already segregated neighborhoods.

The second part of the analysis suggests that the school's demographics do have some significance as predictors of MCAS achievement, even when controlling for non-demographic school characteristics (such as student-teacher ratio or school type). For instance, the proportion of Asian students was significant in all models; other demographic characteristics were significant in some but not others. This finding gives a concrete importance to Boston's policy decisions regarding school assignment and diversity and suggests that the city should investigate the cause of racial discrepancies in schools. The difference in performance between charters and traditional schools also suggests that the city should conduct further studies to determine what features of local charter schools contribute to this high relative performance.

Technical Appendix

We used a chi-squared test to compare the observed number of white students in neighborhoods' schools to the number of white students expected in those neighborhoods based on their proportions of white residents. We conducted three tests, first including all schools, then only traditional schools, and finally only charters. As shown in Table 2, all three tests gave significant

[6] The proportion of Hispanic students in the school was significant at the .05 level, not just the .10 level.

results, providing extremely strong evidence that we can reject the null hypothesis that the proportion of white students matches the proportion of white residents in all neighborhoods.

Table 2

Schools Included	Observations (# of Neighborhoods)	Degrees of Freedom	Test Statistic	*p*-value*
All Schools	17	16	10,642.78	0
Traditional Only	16	15	9,249.072	0
Charter Only	10	9	1,917.13	0

**Each p-value is actually non-zero, but is so small that R does not report a non-zero value.*

We then conducted individual hypothesis tests for each neighborhood, comparing the aggregate proportion of white students in the neighborhood's schools to its proportion of white residents. We included only traditional schools in this analysis, as charter and exam schools do not rely on Boston's usual location-based school assignment system. We used a two-proportion z-test for each neighborhood, as the sample sizes at issue were well above those needed for the Central Limit Theorem to apply. As there were 16 neighborhoods with traditional schools, and thus 16 hypothesis tests to perform, we used the Bonferroni correction to reduce the risk of Type I error across the tests. Thus, we used a .05/16 = .003125 significance level for each test. The results are summarized below. As Table 3 shows, at that significance level, the *p*-value for each test is low enough that we can reject the null hypothesis that there is no difference between the proportions.

Table 3

Neighborhood	White Student Prop.	Number of Students	White Resident Prop.	Number of Residents	Test Statistic	*p*-value*
Allston	0.11	1,242	0.57	19,761	-48.4	0
Back Bay	0.07	389	0.76	17,577	-50.4	0
Brighton	0.15	2,339	0.67	47,768	-66.1	0
Charlestown	0.17	2,465	0.70	18,058	-63.1	0
Dorchester	0.07	10,957	0.22	124,489	-53.0	0
East Boston	0.13	4,951	0.32	44,989	-34.9	$6.20 \cdot 10^{-267}$
Hyde Park	0.07	3,033	0.26	35,585	-35.7	$7.77 \cdot 10^{-279}$
Jamaica Plain	0.15	2,841	0.54	39,240	-55.5	0
Mattapan	0.01	1,614	0.06	24,268	-15.0	$4.64 \cdot 10^{-51}$
Mission Hill	0.06	948	0.47	16,700	-47.9	0
North End	0.53	453	0.91	9,107	-15.9	$5.42 \cdot 10^{-57}$
Roslindale	0.19	2,097	0.51	28,644	-34.6	$9.40 \cdot 10^{-262}$
Roxbury	0.04	5,422	0.11	51,252	-24.1	$2.66 \cdot 10^{-128}$
South Boston	0.22	2,233	0.78	35,660	-61.3	0
South End	0.07	2,610	0.55	31,601	-81.4	0
West Roxbury	0.29	2,667	0.73	32,795	-49.1	0

**Each p-value is actually non-zero, but is often so small that R does not report a non-zero value.*

The second part of the analysis employed multivariate ordinary least-squares regressions to

predict schools' mean English and math MCAS scores based on a variety of school and, in the two larger models, neighborhood characteristics. Schools with fewer than 30 students taking the relevant test type were omitted from the analysis. Initial scatter plots of the data suggested a roughly linear pattern, and plots of the residuals from the models vs. the actual values did not include a discernible pattern, suggesting linear regression was an appropriate method.

The racial proportions used here were selected as covariates because they made up non-negligible proportions of at least some of the schools in the data set and were not very strongly correlated with one another. The analysis excludes the proportion of black students within a school because it correlated strongly and negatively with both the proportion of white students and the proportion of Hispanic students. The full results are displayed in Table 4 below.

Table 4	Dependent Variable: Mean Math MCAS Score		Dependent Variable: Mean English MCAS Score	
Covariate	Smaller Model	Larger Model	Smaller Model	Larger Model
Exam School[a]	-4.56 (4.58)	-6.11 (4.69)	2.68 (3.42)	1.43 (3.58)
Traditional School[a]	-9.35*** (2.08)	-9.85*** (2.10)	-7.22*** (1.55)	-7.85*** (1.61)
Proportion of White Students in School	14.23* (7.65)	12.16 (11.88)	8.67 (5.71)	10.31 (9.08)
Proportion of Asian Students in School	25.55*** (7.42)	31.52*** (8.49)	14.60*** (5.54)	20.63*** (6.49)
Proportion of Hispanic Students in School	7.58** (3.79)	3.51 (5.56)	2.31 (2.83)	4.04 (4.25)
Student-Teacher Ratio	0.39 (0.27)	0.49* (0.27)	0.14 (0.20)	0.15 (0.21)
Proportion of Economically Disadvantaged Students in School	-10.93 (9.27)	-5.60 (9.55)	-12.11* (6.92)	-9.78 (7.29)
Proportion of White Residents in Neighborhood		4.58 (5.05)		-0.51 (3.85)
Proportion of Hispanic Residents in Neighborhood		7.61 (7.99)		-3.00 (6.10)
Proportion of Asian Residents in Neighborhood		-30.44* (16.12)		-24.36* (12.32)
Ratio of Proportion of White Students in School to Proportion of White Residents in Neighborhood		0.01 (0.02)		0.00 (0.02)
Intercept	237.874*** (5.946)	234.50*** (6.42)	243.492*** (4.436)	243.79*** (4.90)
Observations	102	102	102	102
R^2	0.41	0.47	0.51	0.53
Adjusted R^2	0.37	0.40	0.47	0.47

*Standard errors in parentheses. Significance level symbols: * p<0.10, ** p<0.05, *** p<0.01.*
All proportion variables were expressed in decimal form.
[a] The omitted category is Charter School.