Teacher Demographics and Dropout Trends in MA

Project 4: Final Project

Introduction

Throughout this semester, we have been evolving our understanding of the relationship between the percentage of non-white teachers and student dropout rates in Massachusetts high schools. There is a growing disparity between public school teacher demographics and growing diversity of the student body (DESE, 2023). We wanted to know more about this relationship, given that the shifting demographics of the student body in relation to the teachers may influence student engagement, school culture, and overall dropout patterns (Najarro, 2023). Our first analysis was rooted in the principle that representation was important for student success, therefore, schools with a higher percentage of teachers of color, the lower the dropout rates for all students (Egalite, 2024). We found a statistically significant positive association, schools with more non-white teachers tended to have higher dropout rates with a correlation of 0.34 and a fitted slope of 0.14 ($t(401) = 7.25$, $p < .001$).

To gain further insight, we used the existing literature to select relevant variables to examine our relationship of interest using a multiple regression model. Our analysis explored the relationship between the percentage of teachers of color and high school student dropout rates when controlling for the percentage of students who are low-income and the student-teacher ratio. This resulted in a change in our model predictions, indicating that socioeconomic status rather than teacher demographics play a larger role in high school dropout trends. Given the conclusions from previous analyses, our current analysis is focused on charter schools and district schools. Our reasoning for focusing on district and charter schools is rooted in our understanding of education policy debates, where charter and district schools are frequently compared evaluating for equity, student outcomes, and accountability. Since socioeconomic status appears to play a larger role in high school dropout trends, we were curious about the association between teacher demographics and dropout rates depending on school type. Our interest is reflected in our updated research question: **Does the association between the percentage of non-white teachers and student dropout rates vary by school type when controlling for low-income students?**

To answer this question, we use the following population model:

$$student\_dropout_i = \beta_0 + \beta_1(nonwhite\_teachers_i) + \beta_2(low\_income\_pct_i) + \ldots$$

$$\ldots + B_4(typcharter)_i + (\beta_5 nonwhite\_teachers\ x\ charter\_school_i) + \epsilon_i$$

Where student_dropout$_i$ is the dropout rate at school i, nonwhite_teachers$_i$ is the percentage of nonwhite teachers, typecharter$_i$ is an indicator for charter schools (with district schools as the reference category), and low_income_pct_$_i$ is the percentage of low-income students. The interaction term allows the association between teacher demographics and dropout rates to differ by school type.

Data and Measures

The dataset is the Massachusetts Department of Elementary and Secondary Education school-level dataset (MADESE). The analytic sample is restricted to high schools and excludes vocational, virtual, and agricultural schools. Our outcome measure is dropout_pct (dropout rate percentage). The key predictor is nonwhite_teachers, constructed as the sum of the percentages of teachers who identify as Black, Asian, Hispanic, Native American, Pacific Islander, or multiracial. School type is a dichotomous indicator (District vs Charter, with District as the reference). The primary control variable is low_income_pct (percent low-income). Schools with missing values on the outcome or key predictors were removed using drop_na().

Univariate Summaries

After removing missing values on the outcome and key predictors, the analytic sample includes 369 high schools (319 district schools and 50 charter schools). Dropout rates are right-skewed: the median dropout rate is 1.0 percent and the mean is 3.941 percent, with a maximum of 43.7 percent. The percentage of nonwhite teachers is also right-skewed: the median is 6.235 percent and the mean is 14.674 percent, with a maximum of 100 percent. The percentage of low-income students ranges from 2 to 96 percent, with a mean of 33.92 percent. These distributions indicate substantial heterogeneity across Massachusetts high schools in both demographic composition and dropout outcomes. Additionally, it highlights the limited diversity in the greater Massachusetts teacher workforce. Although we are talking

about a small proportion of the population, we are still able to collect evidence of an association between the two variables.

Simple Regression Model

We decided to build on the simple regression model from our previous project, which predicted the association between the percentage of non-white teachers and student dropout rates. The fitted simple model was, $\hat{student\_dropout}_i = 1.84 + (0.14)\ (nonwhite\_teachers_i)$ predicting that if the percentage of non-white teachers is 0, the student dropout rate would be 1.84. The model was statistically significant (t (367) = 6.96, p < .001) and had a positive association between non-white teachers and dropout rates (see Appendix). Moreover, for each one-percent-point difference in the percentage of non-white teachers, there is a 0.14-percentage point difference in the student dropout rate. With an RSME of 7.31 and an R^2 value of 0.12, the model had limited predictive accuracy. This led us to add control variables to better understand dropout patterns. To address this limitation, we controlled for the percentage of low-income students to create our first multiple regression model.

Multiple Regression Model with Controls

The multiple regression model we used was:

$\hat{student\_dropout}_i = -2.68 + (-0.01)(nonwhite\_teachers_i) + (0.20)(low\_income\_pct)$ ...

...+ $(-1.17)(typeCharter)$. In all multiple regression models, district schools are used as the reference category for school type. The intercept of the model changed from 1.84 to -2.68, which means that when non-white teachers, low-income percentage, and school type are 0 (referencing district schools), the predicted dropout rate is -2.68. Although a negative dropout rate is not realistic, this intercept still has meaning. It shows how when we control for low-income percentage and school type, the baseline prediction changes by accounting for differences in socioeconomic status and district or charter schools. The slope for non-white teachers also changed from 0.14 to -0.01. This new slope tells us that after

controlling for socioeconomic status, each one-percent-point difference of non-white teachers is associated with a 0.01-point decrease in dropout rate.

This relationship found in our simple model changed from positive to negative and is no longer statistically significant (t (df = 365) -0.37, p = 0.70). In the case of our model, a negative association is better since it would mean a "decrease" in dropout rates rather than an increase. The change in slope provides further evidence that the simple model's positive relationship overestimated the relationship because it did not account for contextual factors, like the percentage of low-income students. The RSME from the model is 6.27, meaning that the standard deviation of the residuals is around 6.27 percentage points. The model explains about 35% of the variability in student dropout rates (R^2 = 0.35) with the remaining 65% unpredicted. The coefficient for non-white teachers has a 95% confidence interval of [0.03, - 0.05], which includes 0, making us unable to reject the null hypothesis. However, the f-test for the model is significant (f (df = 365) 66.52, p < .001), showing that the model as a whole explains a significant portion of the variation in dropout rates, even though the predictor of non-white teachers is not statistically significant.

Multiple Regression Model with Controls and Interaction

In this model, student dropout rates are predicted by the percentage of nonwhite teachers, the percentage of low-income students, and whether the school is a charter school, including an interaction between teacher demographics and school type. We decided to include the interaction with school type to test for differential effects across district and charter schools. The fitted model is:

$$\hat{student\_dropout}_i = -2.54 + (-0.03)(nonwhite\_teachers_i) + (0.21)(low\_income\_pct_i) \ldots$$

$$+ (-4.05)(typeCharter_i) + (0.11)(nonwhite\_teachers_i \, x \, typeCharter_i)$$

This means that the fitted model for district (public) schools is

$$\hat{student\_dropout}_i = -2.54 + (-0.03)(nonwhite\_teachers_i) + (0.21)(low\_income\_pct_i)$$

while the model for charter schools is

$$\hat{student\_dropout}_i = -6.59 + (0.08)(nonwhite\_teachers_i) + (0.21)(low\_income\_pct_i).$$

The interaction captures the difference in slopes between district and charter schools. In this model, district schools have a slope of -0.03 for nonwhite teachers, while charter schools have a slope of 0.08. The association between nonwhite teachers and dropout rates is 0.11 points more positive in charter schools than district schools. When we are controlling for the percentage of low-income students, a higher share of nonwhite teachers is associated with a slightly lower dropout rate in district schools but slightly higher dropout rate in charter schools. The interaction does not tell us that a type of school is "worse", it only tells us that the strength and direction of the association differs by school type. In this model, the interaction is statistically significant ($p < 0.05$).

The f-test for the model is significant (f (df = 365) 66.84, $p < .001$), showing that the model as a whole explains a significant portion of the variation in dropout rates, even though the predictor of non-white teachers is not statistically significant.

Comparing All Three Models

Across all three models, a positive coefficient is "bad" since it would be associated with a higher predicted dropout rate, while a negative coefficient is associated with a lower dropout rate. We find evidence that the estimated association between the percentage of nonwhite teachers and student dropouts changes once we add controls then add an interaction by school type. In Model 1, which includes only nonwhite teachers, each one-percentage-point difference in nonwhite teachers is associated with a 0.14-point difference in dropout rates ($p < .001$). In Model 2, after controlling for the percentage of low-income students and whether the school is charter or district, the association is no longer present. The slope for nonwhite teachers has decreased by 0.15-percentage points to -0.01 and is also no longer statistically significant. Additionally, it means that the initial relationship (Model 1) can be explained by the percentage of low-income students and school type. In Model 3, we added the interaction between nonwhite teachers and school type. In doing so, we found that the association changes depending on the type of school. For district schools, which are the reference group for the model, the estimated slope is

-0.03 while for charter schools the slope is 0.08. The association between nonwhite teachers and dropouts is 0.11 points more positive in charter schools than in district schools.

	Model 1	Model 2	Model 3
(Intercept)	1.84 ***	-2.68 ***	-2.54 ***
	(0.49)	(0.58)	(0.59)
nonwhite_teachers	0.14 ***	-0.01	-0.03
	(0.02)	(0.02)	(0.02)
low_income_pct		0.20 ***	0.21 ***
		(0.02)	(0.02)
typeCharter		-1.17	-4.05 *
		(0.99)	(1.69)
nonwhite_teachers:typeCharter			0.11 *
			(0.05)
R^2	0.12	0.35	0.36
Adj. R^2	0.11	0.35	0.35
Num. obs.	369	369	369

*** p < 0.001; ** p < 0.01; * p < 0.05

Table 1: *Side-By-Side Model Comparison*

Conclusion

Across all three models, the estimated association between the percentage of nonwhite teachers and student dropout rates is highly sensitive to model specification, which is exactly why the control variables and interaction test matter for answering our research question. In the simple regression model (N = 369), a one–percentage-point increase in nonwhite teachers is associated with a 0.143–percentage-point increase in dropout rates ($p < .001$; $R^2 = 0.12$). This statistically significant bivariate pattern, however, does not persist once we account for major structural differences across schools.

After controlling for the percentage of low-income students and school type, the estimated coefficient on nonwhite teachers shifts from 0.14 to −0.01 and is no longer statistically significant ($p = .709$). This change in magnitude and direction is consistent with omitted variable bias in the bivariate model, because low-income concentration is strongly related to dropout and plausibly correlated with staffing patterns. In the controlled model, low-income percentage is the dominant predictor ($b = 0.2036$, p

< .001), and model fit increases substantially ($R^2 = 0.35$). The nested-model F-test comparing Model 1 to Model 2 confirms that adding low-income percentage and school type meaningfully improves explanatory power (f (df = 365) 66.84, p < .001), directly supporting our decision to include these controls.

When we add the theoretically motivated interaction between teacher demographics and school type, we find evidence that the association differs across institutional contexts. In district schools, the estimated slope is −0.03 (p = .299), while in charter schools the implied slope is −0.025 + 0.11 ≈ 0.089. The difference in slopes, captured by the interaction term, is statistically significant (b = 0.114, p = .0359), providing direct evidence for our interaction hypothesis that the relationship between teacher demographics and dropout is not the same in charter and district schools, conditional on low-income percentage. These estimates should not be interpreted causally, but they do suggest that institutional context is relevant when describing demographic associations.

Overall, the results imply that socioeconomic composition is more strongly and consistently associated with dropout rates than teacher demographics alone, and that conclusions based on simple correlations can be misleading. Some important limitations we recognize of our analysis is that our data is cross-sectional and school-level (not student-level), which are key confounders beyond low-income percentages that are not included. Dropout rates are right-skewed with extreme values that may produce heteroskedasticity and influence; something to take under consideration. Future work should test robustness using additional controls (for example, student racial composition, teacher outreach programs, school size, and urbanicity), consider nonlinear functional forms, and, ideally, use longitudinal or student-level data to better assess mechanisms and reduce concerns about unobserved confounding.

Appendix

Figure 1. Percentage of Students Who Dropout of HS

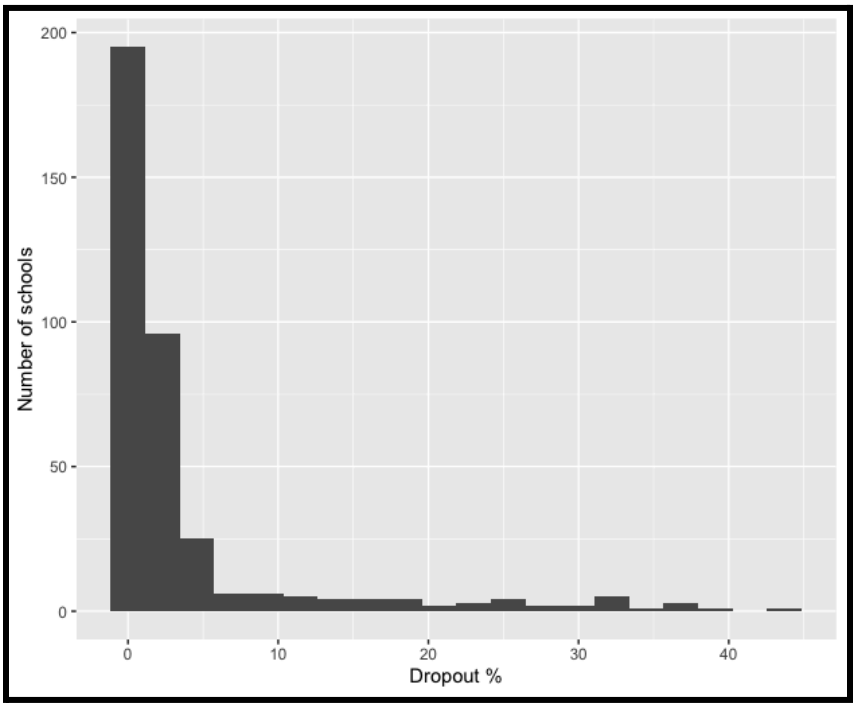


Figure 2. Percentage of Teachers Who Identify as Non-White

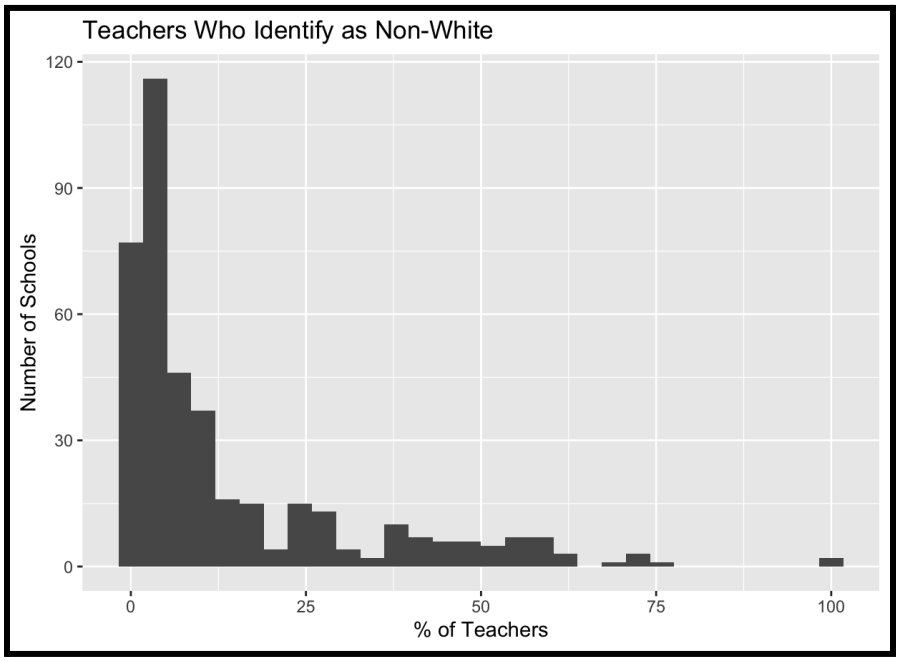


Figure 3. Percentage of Low-Income Students

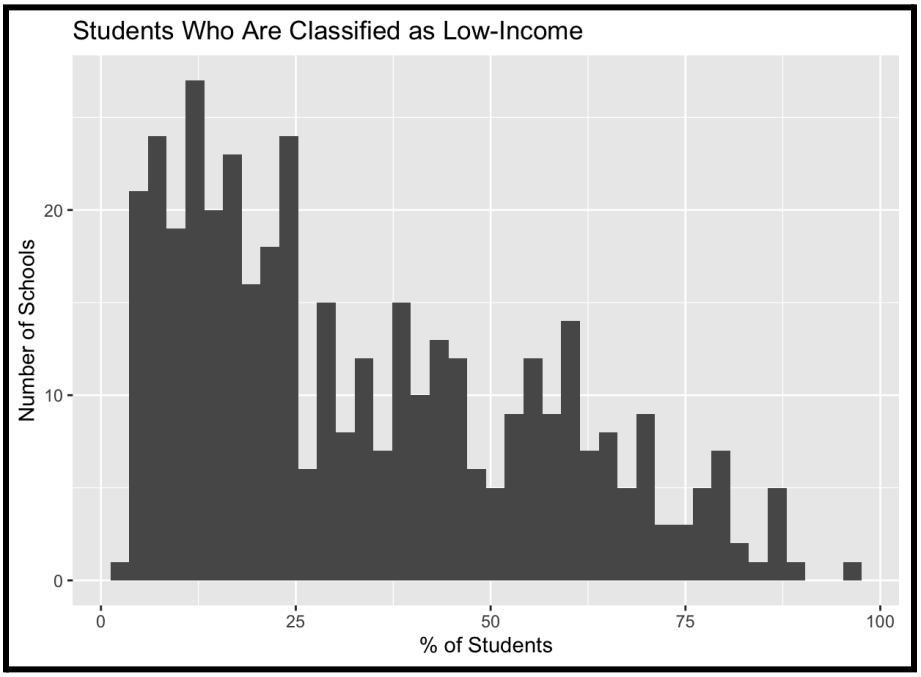


Figure 4. Type of Schools in MA

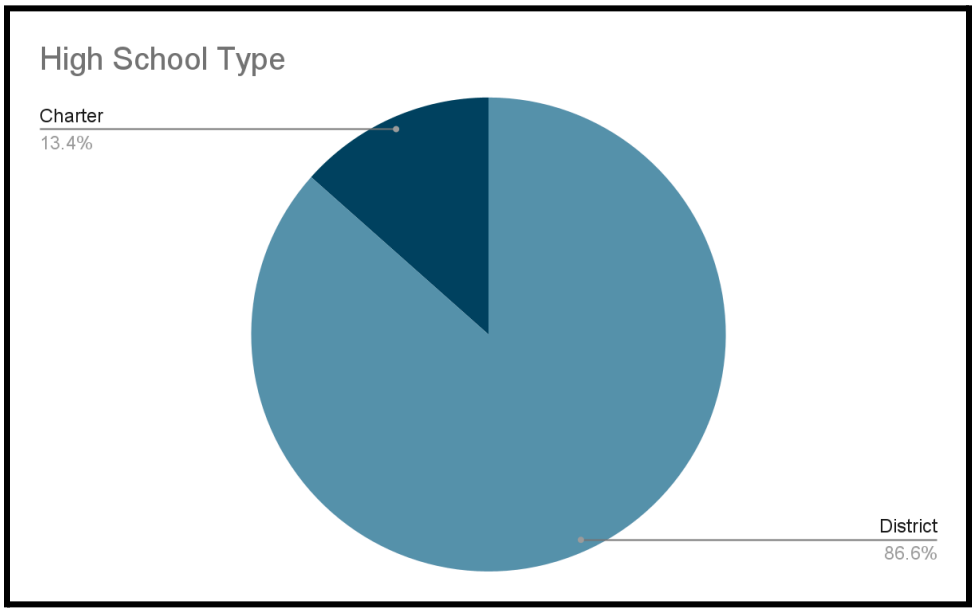


Figure 5. Scatterplot for Student Dropouts and Non-White Teachers

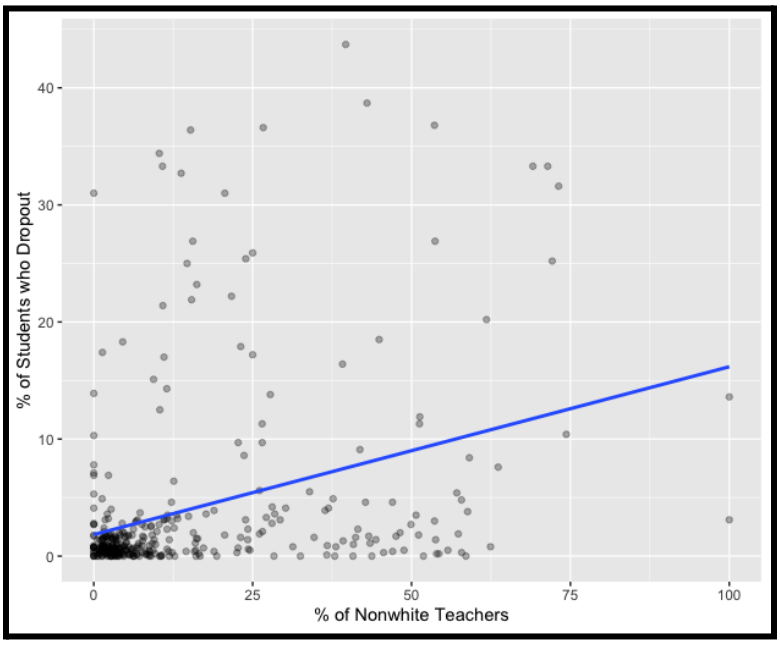
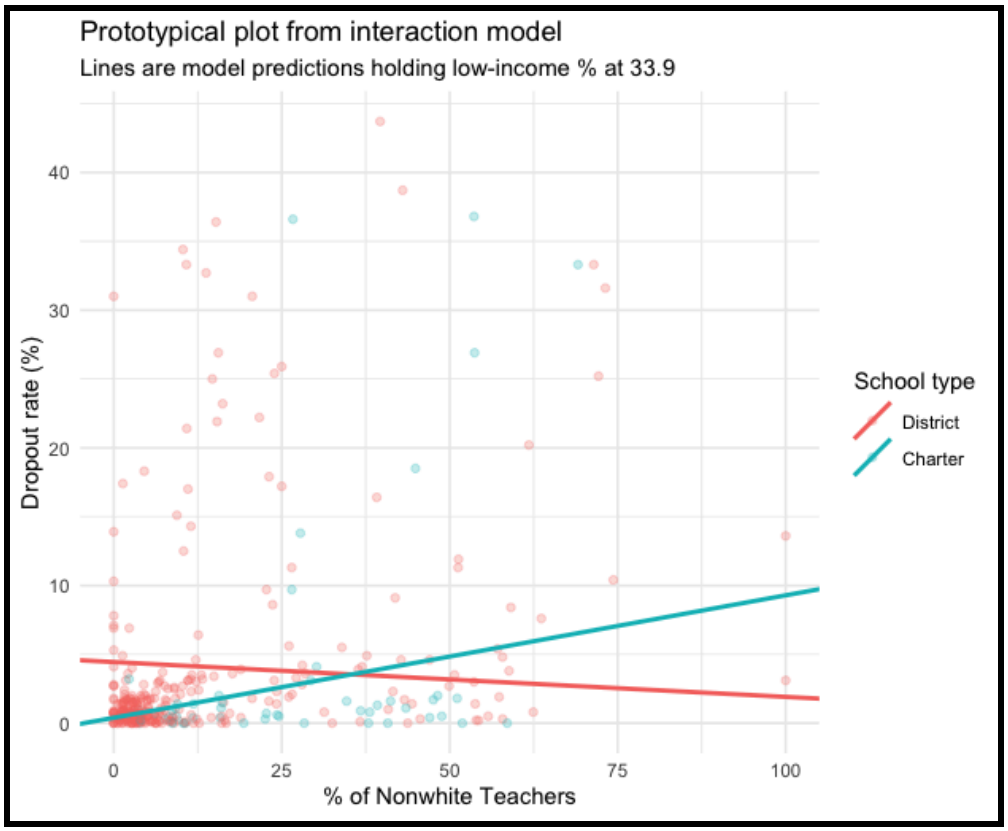


Figure 6. Prototypical Plot



References

Egalite, A. J. (2024). *What we know about teacher race and student outcomes: A review of the evidence to date*. *Education Next*, 24(1), 42-49. https://www.educationnext.org/what-we-know-about-teacher-race-and-student-outcomes/ educationnext.org

Massachusetts Charter Public School Association. (2024, January). *Representation matters: The importance of educator diversity*. Massachusetts Charter Public School Association. Retrieved December 15, 2025, from https://masscharterschools.salsalabs.org/eblastjan_2024

Massachusetts Department of Elementary and Secondary Education. (2023). *The state of educator racial diversity in Massachusetts* [Data story]. E2C Hub. https://educationtocareer.data.mass.gov/stories/s/The-state-of-educator-racial-diversity-in-Massachu/fm6k-958d/

Najarro, I. (2023, February 15). *Districts want more teachers of color but don't do enough to recruit them, survey finds*. Education Week. https://www.edweek.org/leadership/districts-want-more-teachers-of-color-but-dont-do-enough-to-recruit-them-survey-finds/2023/02