

Final Project

Alan Qiao, Jacqueline Dinh

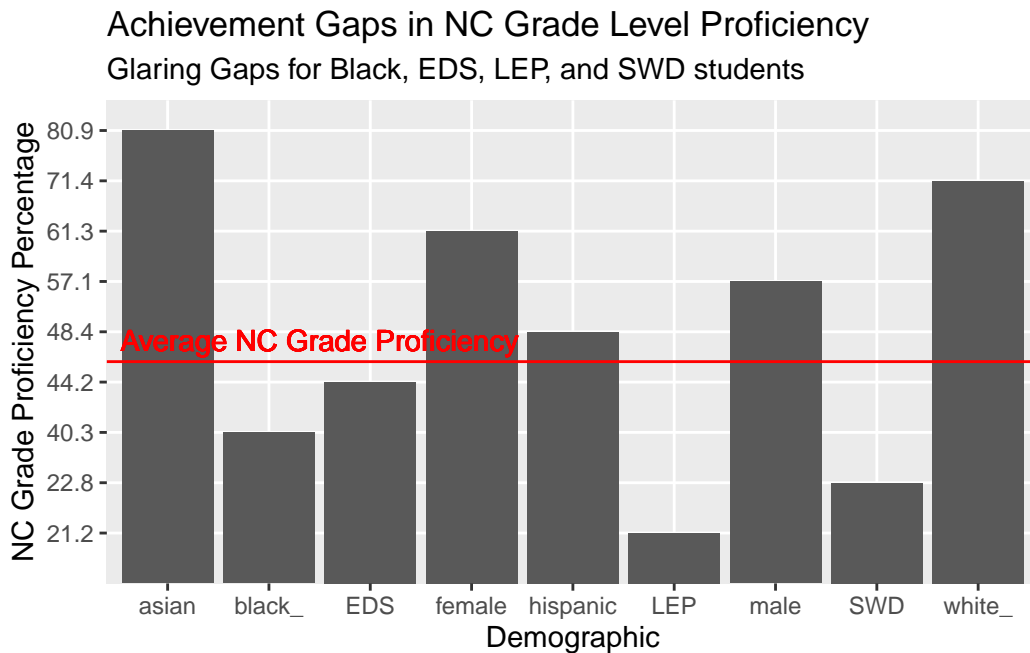
Read in the data

Introduction

As students from the south, where desegregation of public schools moved slowly, with some southern states actively resisting integration (Wilkinson, 1979), we were interested in analyzing how the legacy of segregation has impacted modern-day grade level proficiency in North Carolina public schools. North Carolina, specifically, is seen as a success story for public school integration; in 1971, the historic court case *Swann v. Charlotte-Mecklenburg Board of Education* established a busing system that transported students of color from historically redlined areas of Charlotte to schools located in predominantly white areas. However, in the 90s, the city started to see a pattern of resegregation, similar to that of pre-*Swann* (Mickelson, 2015). In this project, we expanded upon Dr. Roslyn Mickelson's work in *The Cumulative Disadvantages of First- and Second-Generation Segregation for Middle School Achievement*, which found that a student attending a school with a racial imbalance towards a Black population is more likely to be placed in a lower academic 'track', which Charlotte schools used to categorize students based on grade level proficiency.

We wanted to see if the conclusions and implications of Dr. Mickelson's paper rings true in modern times across the entire state of North Carolina. Using a dataset that recorded the percentage of students of different demographics that are grade level proficient and college/career ready in the 2016-2017 academic school year, the bar graph below plots the overall grade level proficiency of students that are Asian, Black, Hispanic, white, are economically disadvantaged, have limited English proficiency, and have disabilities across all North Carolina schools, it can be seen that there is a large gap between white students grade level proficiency (the plotted red line) and Black students, Hispanic students, economically disadvantaged students, students with limited English proficiency, and students with disabilities' grade level proficiency. The bar graph shows there are still existing achievement gaps between minority students and white students. This provided further motivation in analyzing how the percentage of white students and different minority students in a school, shaped by the legacy of segregation, may impact grade level proficiency. Our research question is: how do the white students, students of color, economically disadvantaged students, students with limited English proficiency, and

students with disabilities affect grade level proficiency of a North Carolina public school? We hypothesized that the higher the percentage of white students in a school's student body, the higher the overall grade level proficiency of the school. Conversely, the higher the percentage of students of color, economically disadvantaged students, students with limited English proficiency, and students with disabilities, the lower the grade level proficiency of the school.



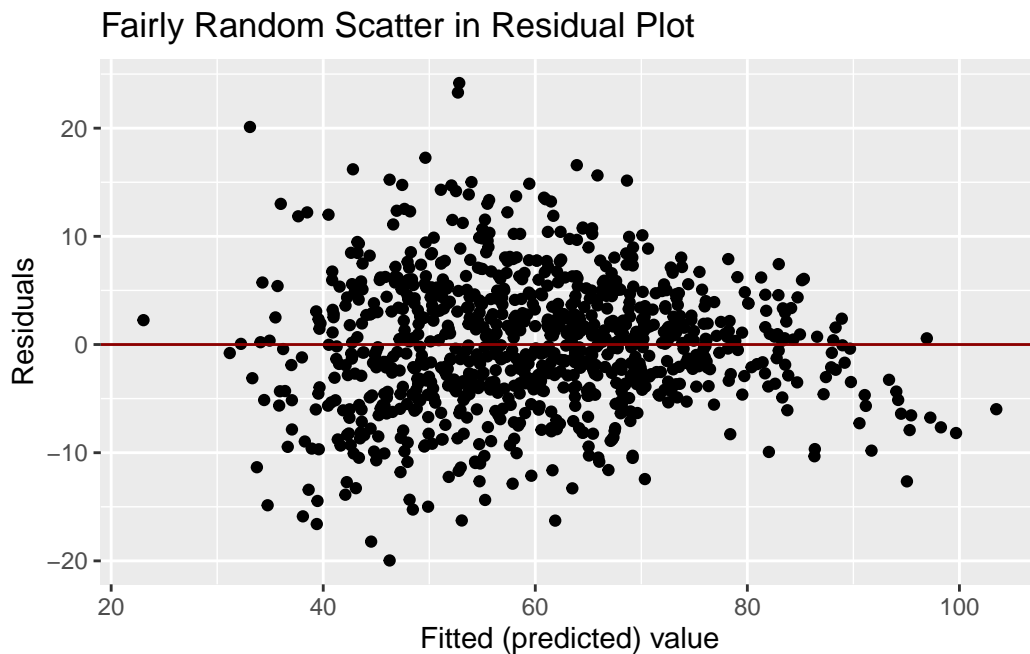
Data

To investigate our research question, we used the dataset NC School Performance (2016 - 2017). This dataset is taken from Durham Open Data, a program managed jointly by the City of Durham's Technology Solutions Department and the County of Durham's Information Services & Technology Department to make data open and free to everyone. Durham Open Data collected this data from the North Carolina Department of Public Instruction and manipulated the annual testing data of public schools in North Carolina from the 2016-2017 school year. The main purpose of this dataset is to record the percentage of students of various races and other factors described that are grade level proficient and college/career ready. In this dataset, there are 32 variables and 85,766 observations separated by school, subject, and standard. The variable Standard was separated into Grade Level Proficient and College and Career Ready, determined by students' test scores of Level 4 & 5 on the CCR for college/career readiness and Level 3 & Above for grade level proficiency. The variable subject included all EOG/EOC Subjects and individual EOC Subjects along with the Math, Science, and Reading EOG and EOC assessments for each grade. To narrow the dataset down and compare schools, we chose

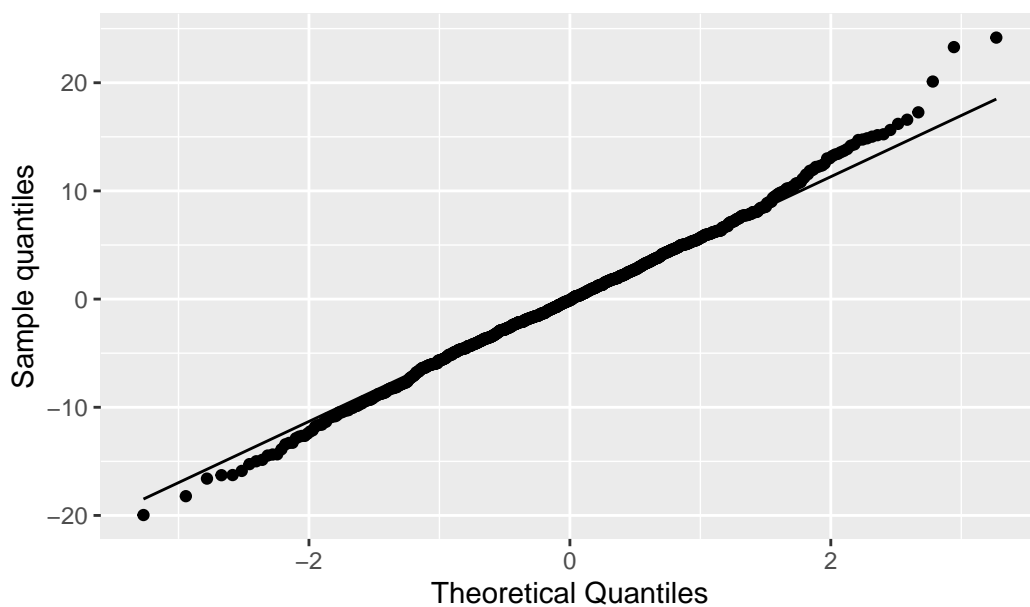
to filter the dataset to view scores of the observations that were determined by the subject of all EOG/EOC Subjects and the standard of Grade Level Proficient.

The dataset presented percentages that are greater than 95 as “>95” and percentages less than 5 as “<5.” To make sure our model does not skip over these observations, we manipulated the dataset to change these percentages to 97.5 and 2.5, respectively, as they are the averages of 95-100 and 0-5. Also, in the dataset, if the number of students that met a condition were less than 10, it would just display “<10.” We then manipulated the dataset again to change these counts to 5, the average of 0 and 10. In the dataset, a * indicates a school does not have tested grades/sufficient data for reporting, and these were changed to NA’s in our model. Rows that did not have sufficient data for the ‘All Students’ subgroup are excluded from this file. In addition, we didn’t include the demographic of American Indian in our model, as there was too little data present in the file because of how small their population is. Finally, we filtered out the State of North Carolina, which was listed as a school, as this was the average of all the schools present and we wanted to investigate individual school demographics.

In our manipulated dataset, to look at and compare an overall assessment for all public NC schools, we included the percentage and number of students that are grade level proficient in all EOG/EOC subjects for all, female, male, American Indian, Asian, Black, two or more races, white, economically disadvantaged (EDS), limited English proficient, disabled, and academically or intellectually gifted students for each school.



Low Deviation for QQ Plot



```
schools <- lm(all ~ female_dem + male_dem + asian_dem + black_dem + hispanic_dem + white_d
summary(schools)
```

Call:

```
lm(formula = all ~ female_dem + male_dem + asian_dem + black_dem +
    hispanic_dem + white_dem + EDS_dem + LEP_dem + SWD_dem +
    AIG_dem + white_comp, data = nc_school)
```

Residuals:

Min	1Q	Median	3Q	Max
-19.9618	-3.8129	-0.0862	3.8170	24.1717

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	251.201098	47.756557	5.260	1.80e-07	***
female_dem	-1.797082	0.489232	-3.673	0.000253	***
male_dem	-1.963939	0.463676	-4.236	2.51e-05	***
asian_dem	0.192906	0.054991	3.508	0.000474	***
black_dem	-0.000600	0.033876	-0.018	0.985874	
hispanic_dem	0.120169	0.040113	2.996	0.002812	**

white_dem	0.169690	0.041335	4.105	4.40e-05	***
EDS_dem	-0.244715	0.015923	-15.369	< 2e-16	***
LEP_dem	-0.053874	0.010088	-5.340	1.17e-07	***
SWD_dem	-0.080386	0.008947	-8.985	< 2e-16	***
AIG_dem	0.506832	0.033728	15.027	< 2e-16	***
white_compbmid white	2.372891	0.809075	2.933	0.003443	**
white_compmminority white	0.185998	1.450784	0.128	0.898015	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.081 on 912 degrees of freedom

(1753 observations deleted due to missingness)

Multiple R-squared: 0.8208, Adjusted R-squared: 0.8184

F-statistic: 348.1 on 12 and 912 DF, p-value: < 2.2e-16

Findings

Based on our model coefficients, it is expected that on average: For every one percent increase of white students in a school's student body the percentage of students that are grade level proficient increases by 0.17%, all else held constant. For every one percent increase of Black students in a school's student body, the percentage of students that are grade level proficient decreases by 0.0006%, all else held constant. For every one percent increase of Asian students in a school's student body, the percentage of students that are grade level proficient increases by 0.19%, all else held constant. For every one percent increase of Hispanic students in a school's student body, the percentage of students that are grade level proficient increases by 0.12%, all else held constant. For every one percent increase of economically disadvantaged students in a school's student body, the percentage of students that are grade level proficient decreases by 0.24%, all else held constant. For every one percent increase of students with limited English proficiency in a school's student body, the percentage of students that are grade level proficient decreases by 0.05%, all else held constant. For every one percent increase of students with disabilities in a school's student body, the percentage of students that are grade level proficient decreases by 0.08%, all else held constant. The expected difference in percentage of grade level proficient students between schools with a white student body majority and a midlevel white student body is 2.37%, all else held constant. This means that the expected difference between a school with a comparatively mid student of color percentage compared to a school with a comparatively low student of color percentage is 2.37%, all else held constant. White minority: The expected difference in percentage of grade level proficient students between schools with a white student body majority and a white student body minority is 0.19%, all else held constant. This means that the expected difference between a school with a comparatively high student of color percentage compared to a school with a comparatively low student of color percentage is 18.60%, all else held constant.

For the predictors for demographics of white students, Asian students, Hispanic students, economically disadvantaged students, and students with limited English proficiency, the p-values are less than our discernibility level of 0.05. Thus, this data does provide convincing evidence that there is a linear relationship between overall grade level proficiency and each of the demographics listed above.

For black students, the p-value is greater than our discernibility level of 0.05, meaning this data does not provide convincing evidence that there is a linear relationship between overall grade level proficiency and black students in a school.

When comparing the levels of percentages of white students and students of color between a school, one of the p-values of the categorical predictor is greater than 0.05. Thus, this data does not provide convincing evidence that there is a linear relationship between overall grade level proficiency and levels of school populations for students of color.

The higher the percentage of white students in a school's student body, the higher the overall grade level proficiency of the school. Conversely, the higher the percentage of students of color, economically disadvantaged students, students with limited English proficiency, and students with disabilities, the lower the grade level proficiency of the school.

Our model supports our hypothesis that the higher the percentage of white students in a school's student body, the higher the overall grade level proficiency of the school. Looking at our results, as the demographic of white students increases, the grade level proficiency of a school increases by 0.17%.

However, when looking at the other individual racial demographics, the model does not support our hypothesis that higher percentages of students of color are associated with lower grade level proficiency in a school. As the percentage of Asian and Hispanic students increase, the grade level proficiency of a school increases. For Black students, we cannot draw conclusions as the relationship between grade level proficiency and Black students are not statistically significant. When looking at the relationship between the grade level proficiency of a school and level of white students, and conversely the students of color, compared across schools of North Carolina, we also cannot draw conclusions as the relationship is not statistically significant.

For students that are economically disadvantaged, students with limited English proficiency, and students with disabilities, our hypothesis is supported as the overall grade level proficiency is expected to decrease with an increase in each of those demographics.

With an adjusted R-squared value of 0.8184, our model 81.84% of variability in the outcome variable can be explained by the model, while adjusting for the number of predictors. This means that our model is considered to be a good fit to predict the grade level proficiency of NC public schools.

Discussion: Association between historical redlining and neighborhood deprivation in North Carolina

As the p-value of the coefficient correlated to white students is higher than 0.05, our hypothesis that the higher the percentage of white students in a school's student body, the higher the overall grade level proficiency of the school is supported. This suggests the legacy of institutional racism and segregation still exist. As people of color, particularly Black individuals, were prohibited access to the same resources and schooling as white Americans, achievement gaps developed between white students and students of color. The model suggests these achievement gaps are present in North Carolina schools. White students are predicted to raise a school's grade level proficiency, depicting the advantage in education white students historically have compared to students of color.

Our model suggested that the same pattern can be found in Asian and Hispanic students. This does not support our hypothesis. However, it raises new and interesting questions that can expand the scope of the research done in this project. *Brown v. Board of Education*, as discussed in the introduction, is typically associated with the integration of Black students into public schools. Nuances between different races are rarely discussed in existing literature. Perhaps there are cultural influences that influence performance of Asian and Hispanic students in North Carolina public schools.

As for the p-values that are not statistically significant, this leads to no conclusion on the relationship between grade level proficiency and Black students and if a school has a minority, moderate, or majority percentage of white students. Our hypothesis is not supported. This calls for a more in-depth analysis of these two specific predictors.

Lastly, the statistically significant p-values for the non-racial minority students—students that are economically disadvantaged, students with limited English proficiency, and students with disabilities—support our initial hypothesis. This supports our hypothesis that an increase in these demographics predicts a decrease in grade level proficiency. Most of our motivation, as discussed in the introduction, centers around racial minorities. Similar to the findings on Asian and Hispanic students, this calls for an expanded research question.

Because we implemented a complete case analysis, some observations were not included in the model, and the model could have not been comprehensive of all schools and their demographics. This could have led to a smaller and less diverse dataset. Since the school year of 2016-2017, much has occurred since then, politically and on a world scale. Because of this, our model may not be representative of how North Carolina schools are faring currently. In addition, this dataset does not include all schools in North Carolina, it only includes public schools. This could have impacted our research question. In the future, we should attempt to gather data from all types of schools in North Carolina to gather a more comprehensive and diverse dataset.

Conclusion: Based on our model, it is predicted that an increase in the percentage of white students, Asian students, and Hispanic students in a school's student population increases the school's grade level proficiency. An increase in the percentage of students that are economically disadvantaged, students with limited English proficiency, and students with disabilities decreases the school's grade level proficiency. There is no relationship between the percentage

of Black students and schools that are either white minority, moderate, or majority and grade level proficiency.

Segregation and institutional racism are present today, but based on our findings with Black, Asian, and Hispanic students, it may manifest in ways not initially hypothesized and thought about in this project. Our findings indicate nuances amongst specific racial minorities. However, notably, non-racial minorities decrease grade level proficiency. For future work, it is important to examine intersectionality between racial minorities and being economically disadvantaged, having limited English proficiency, and having disabilities. Looking at intersections may help further understand how student demographics impact grade level proficiency.

References

Wilkinson, J. H. (1979). From Brown to Bakke: The Supreme Court and school integration: 1945-1978 (pp. 78-128). Oxford Academic.

Jimenez ,M., Lobdell, D.T., Krajewski, A.K., Rappazzo, K.M., Messer L.C., & Luben T.J. (2022). Association between historical redlining and neighborhood deprivation in North Carolina. ISEE Conference Abstracts, 2022(1). <https://ehp.niehs.nih.gov/doi/abs/10.1289/isee.2022.P-0253>

North Carolina Department of Public Instruction (2023). NC School Performance (2016-2017).

Durham Open Data. <https://live-durhamnc.opendata.arcgis.com/documents/4681f6a2614344a496a4bea3e6f8b5>