Impact of Teachers' Level of Education on Student Achievement

Introduction

Over fifty million students attend public schools in the United States and over 190,000 of these students attend public secondary schools in Minnesota. The support and guidance students receive from teachers and faculty in high school is crucial to their academic success. Teachers receive their license and learn how to be effective classroom educators usually through a bachelor's program, but many also go on to receive masters and doctorate degrees. It is thought that a higher level of education is better preparation for any career field, so it is of interest to investigate whether teachers' level of education is impacting their students.

After the No Child Left Behind Act passed in 2001, states and specifically, public schools increasingly began measuring student achievement through state-standardized tests. The Minnesota Comprehensive Assessment (MCA) measures student proficiency rates in mathematics, reading, and science. During high school, the MCA math test is administered to eleventh grade students and proficiency is indicated by meeting or exceeding standards. In this study, we seek to investigate whether higher levels of education of teachers reflects improvements in students' academic proficiency. Specifically, we consider whether a teacher with an advanced degree versus simply a bachelor's degree is associated with an increase in math proficiency rates of high school students.

Many studies have investigated different indicators of student's proficiency rates including economic and demographic factors. The Star Tribune performed a linear regression analysis that compared school's proficiency rates to an expected proficiency rate based on percentage of students on free/reduced lunch, which served as a substitute for income level, in

Minnesota public schools (Appendix A). The achievement gap has also been studied closely by the National Assessment of Educational Progress, which concluded that the white-black gap in reading scores in 2007 was 30 points on the National 12th Grade Assessment (Appendix A).

A similar study performed by Jamie Dial at Baker University investigated teacher's level of education and its effect on achievement on the Missouri Assessment Program. The study concluded that teacher's level of education was not significant in predicting proficiency rates. The study, however, was conducted on only one urban size school district over the course of two school years and included all type of schools in the district. We speculate that looking at all public high schools in Minnesota and accounting for additional economic and demographic factors will indicate a different conclusion that suggests that teachers' level of education has some influence on students' academic success.

Methods

The data for this study was obtained from the Minnesota Report Card at the Minnesota Department of Education's website. Since the website did not provide a comprehensive list of public secondary schools in Minnesota, we obtained a list of schools from schooldigger.com. We imputed these schools into Minnesota Report Card (Appendix A) in order to obtain the relevant information. For this study, we only look into senior high schools that accommodate students from 9th through 12th or 10th through 12th grades to assess test scores exclusive to high school students (for more information on tests, see Appendix A). According to this guideline, we retrieved 235 high schools in total. Our data set contains the following variables: name of the school, zip code, enrollment in grades (9-12 or 10-12), type of educational system (alternative, charter, magnet, virtual, or conventional), white and nonwhite students (%) and free/reduced priced lunch (%) in 2018, mathematics, reading and science proficiency rates (%) in 2016, graduation rate (%) in 2016, number of students and teachers in 2018, and teachers with bachelors, masters and doctoral degrees (%). We included free/reduced priced lunch as it serves as a substitute for income level and graduation rate as it can be an indicator of student's motivation in school. We also added four variables to this data set to account for additional factors: location of high schools (whether high schools are in the Twin-Cities area or not), rate of students not graduating, student-faculty ratio, and advanced (masters and doctoral) to bachelor's degree ratio. For locations of high schools, we matched each zip code with that of the zip codes considered to be in Minneapolis and St. Paul to assess whether the high schools are located in the Twin-Cities area. City-data.com provided us with the data on which zip code corresponds to Minneapolis and St. Paul (Appendix A). This study uses rate of students not graduating instead of the graduating rate, the reason of which will be explained in the subsequent results section under Variable Overview and Transformations.

Results

Modifying the Data Set

Some portions of public high schools were missing pieces of information regarding at least one of the following variables: white/nonwhite percentage, percentage of receiving free/reduced priced lunch, math proficiency rate, graduation rate, number of students/teachers, and percentage of teachers with bachelors/masters/doctoral degree. These schools, total of thirteen, were removed from the analysis. Some schools were only missing data on reading and science proficiency rates and were kept in the model, as this study only considers math proficiency rates. Additionally, the percentage of teachers with either a bachelor's degree or advanced (masters and doctoral) degree was zero at some schools, rendering the advanced to bachelor's degree ratio to be infinite. These cases, a total of seven, were taken out from the analysis. One high school was flagged as an outlier, which will be discussed later, was also omitted from the model. Three schools were identified as having a 100 percent graduation rate, which made the log-transformed not-graduating rate to be infinite. Rather than removing these three schools, however, we added extremely small values to the non-graduation rate, enabling their log-transformation to not be infinite. The following tables show the summaries of each variable of the 214 observations from the data set after removing all the high schools listed above.

Table 1: Numeric Summaries for Math Proficiency, NonWhite and Free/Reduced Lunch Percentage, Graduation Rate, Number of Students and Teachers, and Percentage of Teachers with Bachelors and Advanced Degrees

	Min	1st Quartile	Median	Mean	3rd Quartile	Max	SD
Math (% Proficient)	0.00	31.95	45.00	42.18	56.83	79.30	20.15
NonWhite (%)	2.60	9.22	18.15	29.27	42.58	100.00	26.23
Lunch (% Free/ Reduced)	5.40	19.07	30.95	34.95	45.17	100.00	21.16
Graduation (%)	8.20	83.30	91.25	82.69	95.10	100.00	21.66
Students (Number)	21.0	264.5	622.0	880.3	1462.0	3237.0	732.63
Teachers (Number)	2.00	18.00	34.0	46.8	72.75	178.00	34.89
Bachelors (%)	8.00	28.43	42.85	44.33	59.08	96.80	19.34
Advanced (% Masters and Doctorate)	3.20	39.52	55.65	54.36	70.0	92.00	19.42

Our dataset contains three factor variables: the location of high schools, the enrollment of grades, and the educational type. The location is divided into two groups: Twin-Cities area and outside the Twin-Cities area, the distinction of which is based on the zip code (Table 2). The enrollment of grades is divided into two groups: grades 9-12 and grades 10-12 (Table 3). The educational level is split into 5 categories: alternative learning school, charter school, magnet school, conventional public school, and virtual school (Table 4). Please note that despite the difference in the type of educational system, they are all broadly categorized as public schools.

Table 2: Number of Public High Schools with respect to the location

	Twin-Cities	Non Twin-CIties
Frequency	27	187

Table 3: Number of Public High Schools with respect to Grades

	9-12	10-12
Frequency	206	8

Table 4: Number of Public High Schools with respect to Educational Style

	Alternative	Charter	Magnet	Public	Virtual
Frequency	16	21	3	173	1

Variable Overview and Transformations

Exploratory plots of the data showed a strong right skewness for the distributions of the percentage of nonwhite students, free/reduced priced lunch, number of students, teachers, and the variable of interest, the advanced to bachelor's degree ratio (Figure 1 on the left: nonwhite students: Appendix B). These variables were log-transformed to better fit the normality assumption before any further analysis proceeded. Exploratory plots also showed a strong left skew of the distributions of the percentage of white students, and the graduation rate (Figure 1 on the right: graduation rate). Since the percentage of non-white students is reflective of that of white students in this study, we decided to use a log-transformed version of the percentage of non-white students. For the graduation rate, we calculated the rate of students not graduating by subtracting the rate from 100 to get a variable with a right skew, which was then log-transformed. These log-transformations ensured an approximately linear relationship between the math proficiency rate and the percentage of non-white students, percentage of free/reduced priced lunch, rate of students not graduating, number of students and teachers, as well as the variable of interest, the advanced to bachelor’s degree ratio.

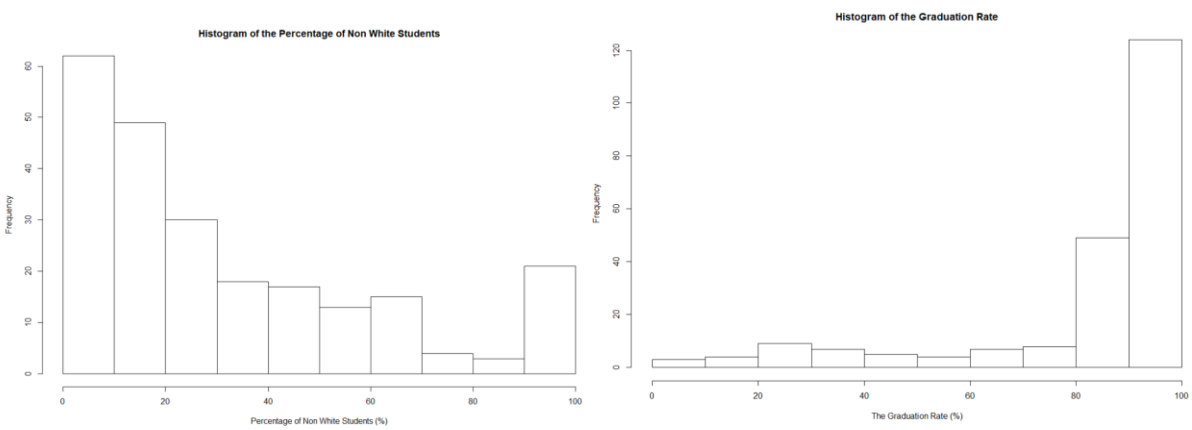


Figure 1: Histograms of the percentage of non-white students (left) and of the graduation rate (right)

Model Assumptions

We assume that all the schools that did not contain the relevant information were omitted at random. Upon creating our linear model, we assume that the observations of each school are independent of each other and that the variance of each variable is constant. When considering a linear model, we also had to ensure the linearity assumption of explanatory variables with math proficiency rates (Appendix C). We also ensured the normality assumption by log-transforming some of the exploratory variables.

Regression Analysis

Scatter plots of math proficiency rates against the log-transformed percentage of non-white students and free/reduced lunch demonstrate a simple linear relationship with possible curvature with math proficiency rates, which will be discussed in the discussion section (Figure 2).

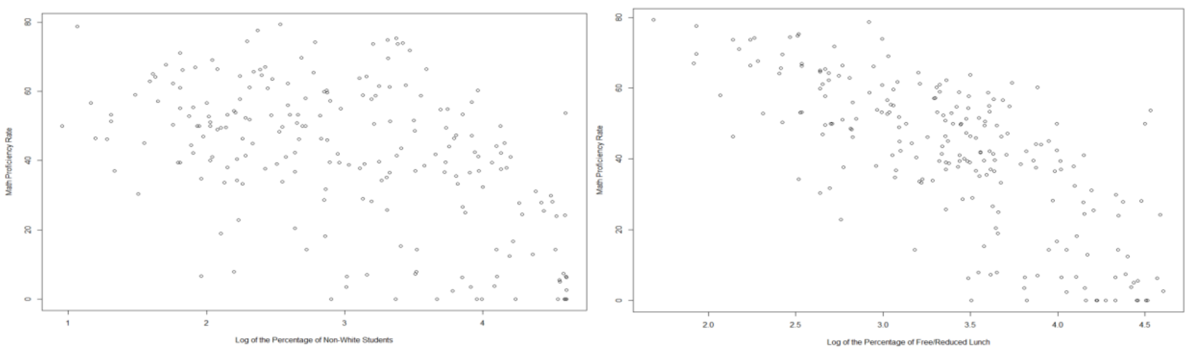


Figure 2: Scatterplots of the Math Proficiency Rate against Log Transformed Percentage of Non-White Students (left) and Log Transformed Percentage of Free/Reduced Lunch (right)

Before we assessed the association between math proficiency rates from the MCA and the advanced to bachelor's degree ratio, we first created a linear regression model that took into account possible confounding variables. Diagnostics during the early stage of analysis found one observation flagged as an outlier through methods such has residuals plots (See Appendix C). This observation, Higher Ground Secondary Academy, is a charter school with a large minority population that is focused on STEM subjects and is thus, not representative of the overall data set. As previously mentioned, this observation was eliminated from the final model. In addition, early models encountered the issue of multicollinearity (See Appendix D for specific values). Number of students and teachers and student-faculty ratio were highly correlated and so, the explanatory variable, number of teachers, was removed from the model to resolve the issue.

Through the analysis, we created a controlled model consisting of five exploratory variables, to which the variable of interest, the advanced-bachelor's degree ratio, was added to see its association with the math proficiency rate. A variety of diagnostics did not detect any issues and the model showed a linear relationship with the response variable (Appendix E). For the final model, we consider the linear model:

$E[\textit{Math Proficiency}] = \beta_0 + \beta_1 log(NonWhite) + \beta_2 log(Lunch) + \beta_3 log(NoGrad) + \beta_4 log(Students) + \beta_5 Type2Alternative + \beta_6 log(DegreeRatio)$

In the above model, Math Proficiency denotes percentage of students meeting or exceeding standards on the MCA, NonWhite denotes percentage of students who are non-white, Lunch denotes percentage of students on free/reduced lunch, NoGrad denotes percentage of students not graduating, Students denotes number of students, Type2Alternative denotes whether the school is type of school (0 = Conventional, Charter, Magnet, or Virtual School, 1= Alternative School), and DegreeRatio denotes ratio of Teachers with Advanced Degree to just a Bachelor's degree.

Table 5: Model Estimates of Parameters

	Estimate	Standard Error	t value	p-value
Intercept	63.944	8.572	7.459	<0.001 (2.35e-12)
Log(NonWhite)	-4.784	1.196	-4.000	<0.001 (8.81e-05)
Log(Lunch)	-11.569	1.865	-6.203	<0.001 (2.97e-09)
Log(NoGrad)	-1.290	0.418	-3.089	<0.005 (2.28e-03)
Log(Students)	5.508	1.018	5.410	<0.001 (1.74e-07)
Type2Alternative	-14.504	3.268	-4.439	<0.001 (1.47e-05)
Log(DegreeRatio)	d2.185	1.009	2.165	<0.05 (3.15e-02)

For a public high school in Minnesota, a doubling of the percentage of students that are non-white is associated with a 3.32 unit decrease in mean math proficiency rates. Similarly, a doubling of the rate of students not graduating is associated with a 0.89 unit decrease in mean math proficiency rates. Even more strikingly, a doubling of the percentage of students with free/reduced lunch is associated with a 8.02 unit decrease in mean math proficiency rates. Surprisingly, a doubling of the number of students at a school is associated with a 3.82 unit increase in mean math proficiency rates at that school. Interestingly, an alternative school is associated with a 14.5 unit decrease in mean math proficiency rates.

We conclude that after accounting for all possible confounding variables, a teacher's level of education is associated with students' math proficiency rates, with a p-value of 0.032. Specifically, a doubling of the ratio (number) of advanced degrees to bachelor's degrees is associated with a 1.515 unit increase in the mean math proficiency rates of students. We are 95% confident that a doubling of the ratio of advanced degrees to bachelor's degrees is associated with a 0.14 to 2.89 unit increase in math proficiency rates.

Discussion

Data from 235 public senior high schools in Minnesota was collected from the Minnesota Report Card on the Minnesota Department of Education. Names of schools along with location (Metro Area or not), percent of student body that is white and using free/reduced lunch, graduation rates, number of students and teachers, and number of teachers with Bachelors, Masters, and Doctorate degrees was also obtained. We used a linear regression model to investigate whether a teacher's level of education is associated with a student's proficiency in mathematics after accounting for school characteristics, demographic, and economic factors. The ratio of teachers with advanced degrees to bachelor's degrees is significant in explaining student proficiency rate (p value = 0.032). It is important to note that while building a controlled model, we observed a curvature between certain exploratory variables and math proficiency rates.

Squared terms were added for the percentage of non-white students and free/reduced priced lunch (both of them log-transformed) to see if they are significant enough to be added in the model. The extra sum-of-square F test, however, showed the squared terms were unnecessary.

The controlled, final model demonstrates a strong association between academic performance of high school students and economic and racial factors, which has also been shown by various scholars. Controlling these influential economic and racial factors, therefore, demonstrates that a teacher's pursuit of a higher education improves math proficiency rates of students of which were measured by the MCA. Our computations demonstrate that a doubling of the degree ratio improves math proficiency rates by 0.14 to 2.89 units. A doubling of the ratio might not be a realistic way to improve a student's academic performance, but in the long run, this could affect hiring decisions in public high schools. Our result differs from the Baker University study as we did find that teachers' level of education was significant. This could, however, be due to the schools in our data set and thus, our model cannot be extrapolated to schools in other states.

While analyzing the dataset, there were several issues that were taken into consideration. First, we removed several schools because they were missing information, information that was probably missing because it was not recorded, although we do not know for sure. Most of the omitted schools were alternative schools, so some of their values can be explained by the school type variable. We hesitate to make the assumption that the missing data was MCAR or MAR. Second, in order to resolve the issue of multicollinearity among the number of students and teachers and student-faculty ratio, we removed the number of teachers from the model. The final model contains, only, the number of students. The positive coefficient of number of students suggests that an increase in number of students at a school improves a student's math test performance. However, this may not be accurate as number of teachers, which was removed, is a confounding factor. In addition, we regarded the percentage of free/reduced priced lunch as an economic factor as it is generally thought to be a substitute for income level since approval for these meals is based on household income. However, this study does not use income level of households directly, which might not fully capture the full significance of economic factors even though the lunch rate already demonstrates an important contribution to the model. Furthermore, the variable that accounts for location of high schools only takes into account whether the school is in Minneapolis or St. Paul and does not account for high schools located in the greater metro area, suburban areas, or rural areas. This slight change in the interpretation of locations might influence the model. Lastly, our data is taken from only Minnesota public high schools and cannot be used to explain proficiency rates of other types of schools nor schools in other states.

Further studies can be conducted using reading and science proficiency rates, which are both variables in the dataset. Comparing across subject can better inform whether a teacher's degree is a good indicator of student proficiency. Additionally, quality of teaching cannot solely be measured by degree obtained, so it would be beneficial to look at number of years of experience of teachers and even whether their master's degree is in education or in another subject. Lastly, future studies can incorporate psychological factors of students such as academic motivation or whether the student prepared for the test. This study might elucidate the importance of soft sides of education rather than hard side of education such as student-faculty ratio. Having this extra information will help school administrators make better hiring decisions and elucidate factors that affect students' academic success.

Appendix

Appendix A: Sources

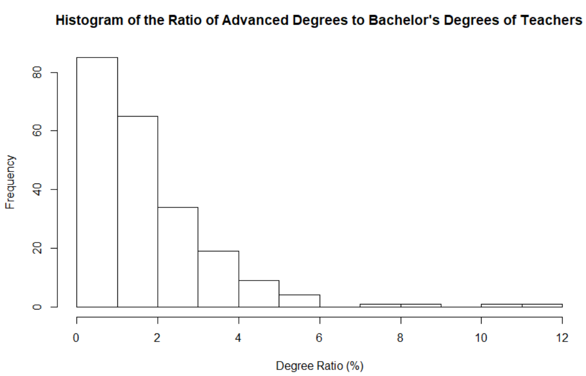
List of Public Secondary Schools in Minnesota:
www.schooldigger.com
Minnesota Report Card
http://rc.education.state.mn.us/#

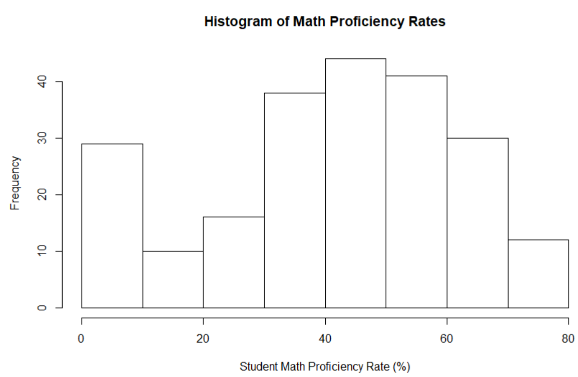
The following sources contain information on zip codes considered to be in the Minneapolis-St.Paul area:
http://www.city-data.com/zipmaps/Minneapolis-Minnesota.html
http://www.city-data.com/zipmaps/St.-Paul-Minnesota.html

The studies cited in this paper were gathered from the following sources:
http://www.startribune.com/find-your-minnesota-school-s-math-and-reading-test-scores/388290262/
https://www.bakeru.edu/images/pdf/SOE/EdD_Theses/Dial_Jaime.pdf
https://nces.ed.gov/pubs2010/2010015.pdf

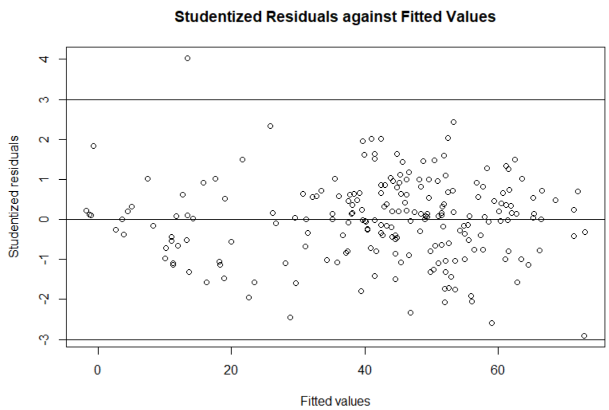
According to the Minnesota Report Card Information Guide, students take mathematics test in grades three through eight and 11, reading in grades three through eight and 10, and science in grades five, eight, and once in high school (during the school year they completed a life science course).
https://education.mn.gov/mdeprod/groups/educ/documents/hiddencontent/bwrl/mdm0/~edisp/mde034431.pdf

Appendix B





Appendix C



Appendix D

	GVIF	Df	GVIF^(1/(2*Df))
logStudents	358.865	1	18.944
logTeachers	284.539	1	16.868
FacultyRatio	20.095	1	4.483

Table 1: Multicollinearity of logStudents, logTeacherse, FacultyRatio

Appendix E

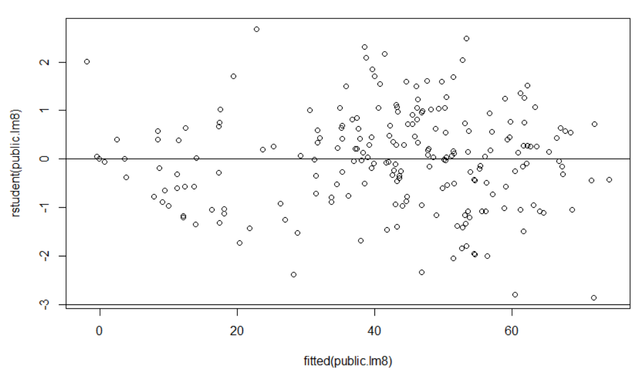


Figure 1: Studentized Residuals Plot of Final Model

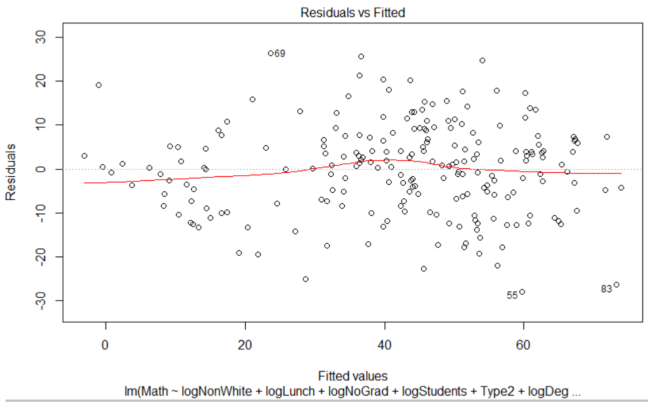


Figure 2: Residuals v. Fitted Values of Final Model

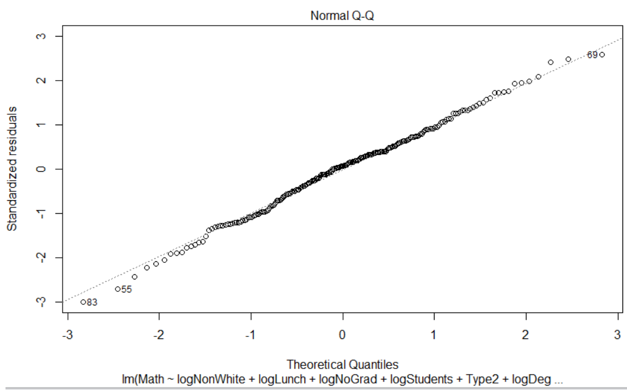


Figure 3: QQ-Norm Plot of Final Model