Unveiling the Recipe for Success: Exploring Key Attributes of High Schools and Their Impact on Graduation Rates

Zoë Schopick

December 2023

1   Abstract

High school graduation rate is an important metric to measure the success of a school. Despite this, little research has been done on the topic, focusing instead on test scores as a measure of student's success. This paper investigates the reasons for the differences seen in graduation rates in public high schools across Virginia. Data was collected for each public high school in Virginia on attributes of the school as well as attributes of the county. A multiple linear regression was performed to create a model for graduation rate as well as student to teacher ratio. Significant predictors were found for each. The explanatory variables were able to explain a relatively high percent of the variation in the response variables. This study leads to the understanding that factors outside of a student's control are playing too high of a role in influencing their chances of graduating high school.

2 Introduction

High school graduation rate is a key way to measure the health, education, and skill level of American society. Completing high school is a key milestone in educational achievement as well as social and economic advancement, "For individuals, a high school diploma has long been recognized as an essential step towards economic and social well-being. Individuals with higher levels of education (and more advanced credentials) enjoy higher income, more stable employment, and less dependency on public assistance" [12]. Graduation rates are an important indicator of the success of a school system. Despite its importance, graduation rates have not been the major focus of education research or statistics, with much of the past research focusing on test scores [12]. Recently, scholars have begun to question whether standardized test scores are a metric that accurately measures student achievement and success [2]. Graduation rate is a more objective metric to measure student's success. The average high school graduation rate in the United States is currently 87%. Within states the graduation rate ranges from 73% in Washington DC to 92% in Iowa and West Virginia [6]. This study focuses on Virginia specifically where the graduation rates of public high schools range from greater than 99% at the highest to around 40% at the lowest [9]. This study seeks to find reasons for such large differences in public school graduation rates.

Past research into graduation rates has mostly been focused on the individual student. Studies have looked at factors such as attendance, passed classes, student engagement, and economic factors to explain whether or not students graduate from high school [10]. My study looks at factors on a school level rather than an individual level. This is important because it indicates which areas schools can improve upon when looking to increase student success. The current study expands upon research done by Swanson [12]. Swanson found

relationships between graduation rate and the racial makeup of schools, the location of schools, and the amount of students receiving free and reduced lunch. Swanson studied these factors in 2001 [12]. I wanted to see if these factors remain impactful to the graduation rates twenty years later and if there were other factors that are impactful as well.

To explore the differences in graduation rates among schools, this paper investigates the relationship between graduation rate and attributes about the schools (Title I status, student to teacher ratio, the number of students eligible to receive free and reduced lunch, and location) and the counties (percent of the county that is white and percent of the county with a college degree). A school is designated as Title I if a certain percent of the student population lives below the poverty lines. These schools receive federal funding for programs that work with failing and struggling students, "Title I, Part A provides financial assistance through state educational agencies to school divisions and public schools with high numbers or percentages of children from low-income families to help ensure that all children meet challenging state academic content and achievement standards" [16]. Student to teacher ratio is calculated by the number of students in the school divided by the number of teachers in the school. It can be understood as the average number of students there are per one teacher in the school, "The pupil/teacher ratio measures the number of students per teacher. It reflects teacher workload and the availability of teachers' services to their students. The lower the pupil/ teacher ratio, the higher the availability of teacher services to students. The pupil/teacher ratio is not the same as class size" [7]. Students are eligible to receive free and reduced lunch based on household size, household income, and homelessness status, "children who are members of households receiving Supplemental Nutrition Assistance Program (SNAP) benefits (formerly the Food Stamp Program) or who receive Temporary Assistance

for Needy Families (TANF) may be automatically eligible for free meals" [15]. This research searches for any significant correlations between the graduation rate and the attributes of schools and counties as well as variables that explain the wide variation seen between the graduation rates in each school.

The data for this project was compiled from the census, the Virginia Department of Education, and the National Center for Education Statistics. Variables included in this dataset are: graduation rate, location, Title I status, number of students, student to teacher ratio, free and reduced lunch, percent of the county with a college degree, and the percent of the county that is white. Each variable was collected once for each public school in Virginia. A multiple linear regression was used to create one model for graduation rate and one model for student to teacher ratio. These models were used to explore the variance in the graduation rate across schools.

There were significant relationships found for both graduation rate and student to teacher ratio. The adjusted R squared values were low, but significant in context of the model. Title I status, student to teacher ratio, and percentage of the county with a college degree were found to be significant predictors for graduation rate. The amount of students receiving free and reduced lunch and the percent of the county that is white were found to be significant predictors for the student to teacher ratio.

The models' ability to account for around a fifth of the variation seen in the graduation rate and student to teacher ratio helps us to gain important insight into the public school system. In a perfect world, graduation rates would be equal for all schools. Students should have an equal chance in graduating, if they put in equal effort, no matter what school they go to. Based on the research, this is not the case. These findings add to an ongoing conversation in college admissions about how to fairly evaluate and compare different students [18].

This research suggests that not all high schools are equal and so the metrics on which students are evaluated for college and careers need to be adjusted for each individual student.

In what follows, I first introduce the data and how it was collected. Then I explain the methods used to analyze the data and the models that were created. An adjusted R squared values table, a correlation matrix, and graphs for each of the response variables are presented. The implications of this research are then discussed, and finally possible future research is presented.

3   Data and Methods

The dataset used for this project was compiled from three different sources. Census data provided information on a county level. The data included from the census was the population of each county, the percentage of the county that is white, the percentage of adults who graduated high school, and the percentage of adults who graduated college [13]. The National Center for Education Statistics provided data on each individual high school in Virginia. Data was included for each public school in Virginia. The factors that were included are the school identification number, the school name, the county name, the city name, what type of location the school was in (city/suburb/town/rural), Title I status of the school, the number of students, the student to teacher ratio, and the number of students who were eligible to receive free and reduced lunch [7]. The final data source was the Virginia Department of Education which provided the number of students who graduated from each high school [17]. These three datasets were merged together to create the final dataset used for analysis. Schools were only included if they were currently open. The data was manually checked to catch any duplicate schools that were referred to by different names. The data of these schools were combined. In the final dataset the variables included were - county name, school name, location, number of students, Title I status, student to teacher ratio, the number of students eligible to receive free and reduced lunch, the percentage of the county with a college degree, the percentage of the county that is white, and the number of students who had graduated. Free and reduced lunch was converted to a percentage based on the total number of students. Graduation rate was converted to a percentage based on the total number of students (the number of seniors or grade twelve students was not used here because accurate data could not be found). The location of the school was changed to numeric with the schools that were the most rural (Rural: Remote)

being coded as a twelve and the schools that were the most urban (City: Large) being coded as a one. The other locations were coded as numbers between one and twelve. School and county names were removed before analysis. Schools with missing data were removed before analysis.

The dependent variable is the graduation rate for each school. This specific measure was chosen because it has been recognized as a way to study the success of schools [12]. Based on the analysis performed, student to teacher ratio was also used as a dependent variable. This variable was used because it was found to be a significant predictor to graduation rate and there has been little research done on what impacts the student to teacher ratio.

The independent variables are the attributes of the individual high schools and counties. They are - number of students, Title I status, student to teacher ratio, percentage of students eligible to receive free and reduced lunch, the percentage of the county with a college degree, and the percentage of the county that is white. For the analysis of student to teacher ratio, all of the same independent variables were used, but student to teacher ratio was removed as well as any factors that were significant predictors for the graduation rate. These aspects were chosen to study based on past research. Student to teacher ratio has been shown to be associated with higher graduation rates [11]. Swanson [12] found relationships between graduation rate and the racial makeup of schools, the location of schools, and the amount of students receiving free and reduced lunch. His study was done in 2001 so I wanted to investigate these same factors so see if they were significant twenty years later. Title I schools are given more federal funding because of high poverty levels [14]. I wanted to investigate if this increased funding impacted graduation rates. There have been studies that have investigated the relationship between parents' education levels and the educational success of the child [4]. I wanted to see if this was similar on a

8

county level.

A correlation analysis was performed on the data to see if there were any strongly correlated factors. A multiple linear regression was performed twice to create two different models. Models were created to predict the graduation rate of high schools and the student to teacher ratio of high schools. All other variables (location, Title I status, percentage of students eligible to receive free and reduced lunch, the percentage of the county with a college degree, and the percentage of the county that is white) were used as predictors (student to teacher ratio was also used as a predictor for graduation rate). A multiple linear regression was chosen because the goal was to be able to explain the differences of these response variables based on multiple quantitative, independent factors.

4 Results

Graduation Rate		
	Estimate	P-Value
Title I	-3.23	0.01
Student/Teacher Ratio	-0.53	<0.001
Bachelor's Degree	0.04	0.01

Table 1: Linear Model of Graduation Rates

Table 1 shows the model created for graduation rate. Title I status, student to teacher ratio, and the percentage of the county with a Bachelor's degree were all found to be significant predictors for graduation rate. The estimate gives the expected change in the response variable due to a change in that explanatory variable. Based on the estimates, Title I status has the most impact on the graduation rate followed by student to teacher ratio and then Bachelor's degree percentage.
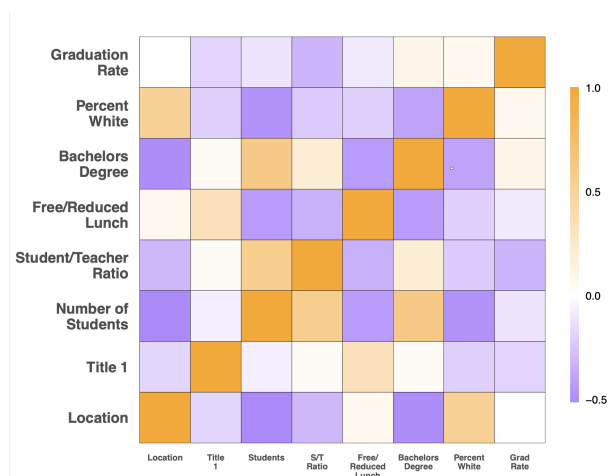

Figure 1: Correlation Matrix

Figure 1 shows a correlation matrix. The correlation matrix shows the relationships between each variable in the dataset. The variables with highly saturated color (either purple or orange) are closely correlated either positively or negatively. The bright orange squares along the diagonal have the highest possible correlation of one because they represent the correlation between two of the same variable.

Adjusted R Squared Values	
Graduation Rate	Student to Teacher Ratio
0.15	0.20

Table 2: Adjusted R^2 values for each model

Table 2 shows the adjusted R squared values for each model created. The first row names the model by its response variable. The adjusted R squared value identifies the percentage of variance of the response variable that is explained by the explanatory variables. Different mixes of the explanatory variables were used to create each model. Each graph shown below is created with the explanatory variables that were used to create the best model for each response variable. The model with the lowest adjusted R squared value was chosen as the best model for each response variable.

Figure 2 shows a violin plot comparing the graduation rates of Title I schools and non Title I schools. One shape is shown for each status on the x-axis. The y-axis shows the graduation rates based on the population of the entire school. The plots are wider where there are more data points and thinner where there are less. The Title I schools have the majority of data points centered around 20% and the non Title I schools have the majority of the data points centered just below 25%. Title I schools have a significantly lower graduation rate than non Title I schools.
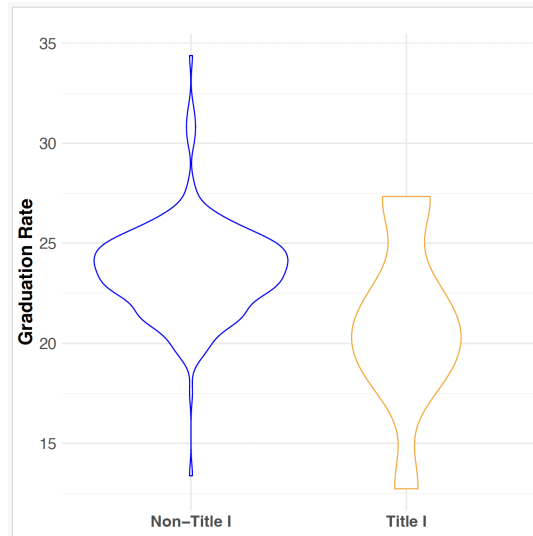
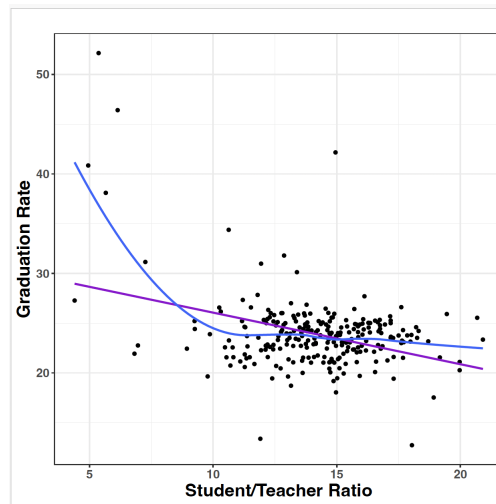Figure 2: Title I Status and Graduation Rate

Figure 3: Student to Teacher Ratio and Graduation Rate

Figure 3 shows the relationship between student to teacher ratio and graduation rate. Student to teacher ratio is along the x-axis and graduation rates based on the population of the entire school are on the y-axis. Two trend lines

are drawn. A purple line of best fit is drawn based on a linear equation and a blue line of best fit is drawn based on an exponential equation. The linear equation was used to fit the model, but an exponential equation is shown because it can also be a good fit for the data. One data point is shown for each school. The data points center around the middle, but show a moderately negative linear trend. As the student to teacher ratio increases, the graduation rate decreases.

Student to Teacher Ratio		
	Estimate	P-Value
Free and Reduced Lunch	-0.04	<0.001
Percent White	-0.04	<0.001

Table 3: Linear Model of Student to Teacher Ratio

Table 3 shows the model created for student to teacher ratio. The percentage of students eligible to receive free and reduced lunch as well as the percent of the county that is white were found to be significant predictors for student to teacher ratio. Based on estimates, both of these variables show the same, small, impact on student to teacher ratio.
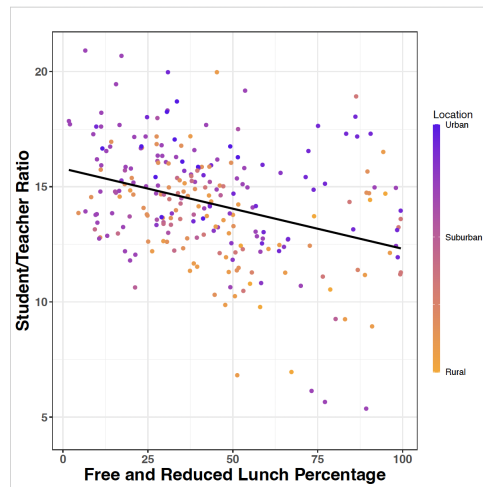

Figure 4: Free/Reduced Lunch and Student to Teacher Ratio

Figure 4 shows the relationship between the percentage of students eligible to receive free and reduced lunch and the student to teacher ratio. There is one data point shown for each school. The data points are colored by location with dark blue being urban, purple being suburban, and orange being rural. The x-axis shows the percentage of students receiving free and reduced lunch. The y-axis shows the student to teacher ratio. There is a line of best fit shown in black created by a linear model. There is a negative correlation between free and reduced lunch percentage and student to teacher ratio. As the number of students eligible to receive free and reduced lunch increases, the student to teacher ratio decreases. The graph also shows orange data points towards the bottom of the graph. Looking at Figure 2 we see there is a negative correlation between location and student to teacher ratio. As schools become more rural, the student to teacher ratio decreases.
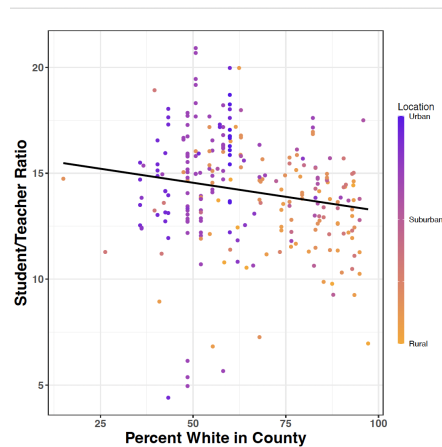

Figure 5: Percent White and Student to Teacher Ratio

Figure 5 shows the relationship between the percent of the county that is white and the student to teacher ratio. There is one data point shown for each school. The data points are colored by location with dark blue being urban,

purple being suburban, and orange being rural. The x-axis shows the percent of the county that is white. The y-axis shows the student to teacher ratio. There is a line of best fit shown in black created by a linear model. There is a negative relationship between percent white and student to teacher ratio. As a county becomes more white, the student to teacher ratio decreases. The graph also shows orange data points predominantly on the right side of the graph. Looking at Figure 2 we can see there is a positive correlation between location and percent of the county that is white. As schools become more rural, the percent of the county that is white increases.

5   Discussion

Multiple significant relationships were found to add to the research on graduation rates in United States public schools. Title I status was found to be a significant predictor of graduation rate. Title I schools have a significantly lower graduation rate than non Title I schools. This indicates that the federal funding that is going towards Title I schools because of the economic situation of their student population may not be working as intended. Title I services are meant to focus on students who are failing or most at risk of failing and bring them back up to academic standards [16]. Based on the lagging graduation rate, these funds may not be enough to make Title I school performances equal to those of non Title I schools. Student to teacher ratio was also found to be a significant predictor of graduation rate. Figures two and five show a negative correlation between student to teacher ratio and graduation rate. As the student to teacher ratio decreases, the graduation rate increases. This research suggests that as there are less teachers per one student in the school, student performance is better. These findings agree with previous studies which also find an increase of the number of students per one teacher results in a decrease in academic performance [8]. The percent of the county with a Bachelor's degree was the final significant predictor of graduation rate. Figure two shows a positive correlation between the number of people in the county with a Bachelor's degree and the graduation rate. As people in a county become more educated, the academic achievements of the students at a school in that county go up. These findings agree with and expand upon past studies on this subject which find a strong positive correlation between a parents' education level and their children's academic achievement [1].

Because student to teacher ratio was found as a significant predictor of graduation rate and there is little research done into the cause of high or low student

to teacher ratios, the study chose to investigate the relationship between student to teacher ratio and attributes of schools and counties. Free and reduced lunch was found to be a significant predictor of student to teacher ratio. Figures two and seven show a negative correlation between free and reduced lunch percentage and student to teacher ratio. As the number of students who are eligible to receive free and reduced lunch increases, the student to teacher ratio decreases. More students who are eligible to receive free and reduced lunch leads to less students for every one teacher in the school. Figure seven also shows that there is a relationship between location and student to teacher ratio. As schools become more rural, the student to teacher ratio decreases. There has not been previous research done on this subject, but the reason for this relationship could be the lack of qualified teachers employed by high-poverty schools. When a large percentage of the student body qualifies for free and reduced lunch it means that the school is high-poverty. These types of schools are more likely than low-poverty schools to hire teachers with few credentials [3]. The lower student to teacher ratio in these high-poverty schools could be due to the lack of credentials. More teachers and staff could be needed because these teachers lack the necessary credentials and experience. The percentage of the county that is white was also found to be a significant predictor of student to teacher ratio. There is a negative relationship between percent white and student to teacher ratio. As a county becomes more white, the number of students for every one teacher decreases. There has not been previous research done on this relationship, but it is probably due to the discrepancies in funding between mostly white districts and districts with predominantly students of color, "school districts that predominantly serve students of color received $23 billion less in funding than mostly white school districts in the United States in 2016, despite serving the same number of students" [5]. When counties have

17

more money, like white counties do, they can hire more teachers and staff.

The adjusted R squared values of the models created shed more light onto the relationships that were discovered. The model for graduation rate had an adjusted R squared value of 0.15. This says that 15% of the variation in the graduation rate was due to the explanatory variables (Title I status, student to teacher ratio, and percent of the county with a college degree). Though this seems low, 15% is relatively high in the context of graduation rate. The models only take into account variables that are outside of each student's control. There are no variables that account for study time, effort put into school, or overall intelligence. The ability for a student to graduate high school should be solely due to the effort the student puts into school. The fact that 15% of variation in graduation rates can be attributed to factors completely outside of a student's control indicates that public schools are not serving their students adequately. Graduation rate should be equal across all schools, no matter the outside factors.

The model for student to teacher ratio had an adjusted R squared value of 0.20. This says that 20% of the variation in student teacher ratio was due to the explanatory variables (free and reduced lunch and the percent of the county that is white). Similar to the model for graduation rate, this low adjusted R squared value is actually high in context. Low student to teacher ratio is best for student learning [11]. Every student deserves to have the best chance in learning. Student to teacher ratio should not depend at all on factors outside of the control of school administrators.

A strength of this study is the validity of the data. All data was collected from government sources from public schools. This data was all from the same year and was accurate in its collection. The data ethics is also a strength of this study. Data about poverty and demographics can be sensitive and easily traced back to the source. If percentages were too small, and could result in

re identification, I did not include them in the data table. This is the reason that percent white alone was included in the study and not the percent of each race/ethnicity. Some of the categories had too little data and could be traced back to the individual students. With the data as it is, each student's identity is protected. Some weaknesses in the data include outliers. There were multiple graduation rates that were calculated that did not seem logical. They were above 50% or even 100% of the entire school population. Graduation rates above 100% were excluded, but no other data points were excluded. These outliers might have been misreported data, but there is no way to tell. The outliers probably impacted the final models. Another limitation of this study was the missing data. Since data was collected from multiple sources, not all schools had the same data, and there was no way to find this missing data. Because of this, many schools had to be excluded from the analysis which created less accurate models.

This research suggests important implications about the United States public high school system. Based on the lower graduation rate of Title I schools compared to non Title I schools, the federal Title I funding may not be working as intended. A relatively high adjusted R squared value for the graduation rate model suggests that outside factors are playing too high of a role in impacting the graduation rate of students. The study suggests that schools, administrations, and governments focus on addressing inequities in schools based on their status, the amount of teachers, and the educational attainment of the county. Understanding the factors influencing graduation rate can help to find where funding and attention should be focused.

This research could be expanded in many ways. Only Virginia high schools were used in this study. Future studies could look at the same factors in other states and see if similar relationships are discovered. The study was also only

done in the 2021-2022 school year. The study could be expanded to other years to see if there are any patterns that appear. There are also other factors that have yet to be explored. Racial and gender make ups of individual schools could be studied as well as the types of classes each school offers.

6   Conclusion

This study sought to find an explanation for the difference in graduation rates for public high schools in Virginia. School location, number of students, Title I status, student to teacher ratio, the number of students eligible to receive free and reduced lunch, the percentage of the county with a Bachelor's degree, and the percentage of the county that is white were used as explanatory variables for the response variable of graduation rate. Each data point was collected for each public high school in Virginia. A multiple linear regression was used for analysis. Title I status, student to teacher ratio, and percent of the county with a Bachelor's degree were found to be significant predictors for graduation rate. Student to teacher ratio was also studied as its own response variable. The percentage of students receiving free and reduced lunch as well as the percent of the county that is white were found to be significant predictors of student to teacher ratio. Relatively high adjusted R squared values indicate that the variation of the response variables were due, in part, to the explanatory variables. These adjusted R square values suggest that outside influences play too high of a role in a student's success in high school.

References

[1] Dr. Muhammad Idris, Dr. Sajjad Hussain, & Dr. Nasir Ahmad. (2020). Relationship between Parents' Education and their children's Academic Achievement. *Journal of Arts & Social Sciences*, 7(2), 82-92. https://doi.org/10.46662/jass-vol7-iss2-2020(82-92)

[2] Goldhaber, D., & Özek, U. (2018, December 6). *How much should we rely on student test achievement as a measure of success?*. Thomas B Fordham Institute. https://fordhaminstitute.org/national/commentary/how-much-should-we-rely-student-test-achievement-measure-success

[3] Hong, J., & Yee, E. (2022, July 21). *Low-Income Students Are More Likely to Be in Classrooms With Underqualified Teachers*. KQED. https://www.kqed.org/news/11920137/low-income-students-are-more-likely-to-be-in-classrooms-with-underqualified-teachers

[4] Lamar University. (2021, August 4). *Correlation Between Parents' Education Level and Children's Success*. Lamar Edu. https://degree.lamar.edu/online-programs/undergraduate/bachelor-science/university-studies/parents-education-level-and-childrens-success/

[5] Mervosh, S. (2019, February 27). *How Much Wealthier Are White School Districts Than Nonwhite Ones*? $23 Billion, Report Says. The New York Times. https://www.nytimes.com/2019/02/27/education/school-districts-funding-white-minorities.html

[6] National Center for Education Statistics. (2015, May). *Teachers and Pupil/Teacher Ratios . School Characteristics and Climate*. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://nces.ed.gov /programs/coe/pdf/Indicator_CLR/coe_clr_2015_05.pdf

[7] National Center for Education Statistics. (2023). *School Search*. Common Core of Data. https://nces.ed.gov/ccd/schoolsearch/school_list.asp?

[8] Nizamettin Koc, Bekir Celik, (2015, April 22). *The Impact of Number of Students per Teacher on Student Achievement*, Procedia - Social and Behavioral Sciences, Volume 177, Pages 65-70, ISSN 1877-0428, https://doi.org/10.1016/j.sbspro.2015.02.335.(https://www.sciencedirect.com /science/article /pii/S1877042815016894)

[9] Public School Review. (2023). *Top 10 Best Graduation Rate Public Schools in Virginia*. https://www.publicschoolreview.com/graduation-rate-stats/virginia

[10] Ritter, B. (2015). *Factors Influencing High School Graduation*. Washington Student Achievement Council . chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj /https://wsac.wa.gov /sites/default/files/2015.12.3.Ritter.Graduation.Issue.Brief.pdf

[11] Spallino, J. (2016, September 28). *Does Class Size Impact Student Achievement?*. Method Modern Schools. https://www.methodschools.org/blog/doesclasssizeimpactstudent-achievement

[12] Swanson, C. (2004, August 31). *The Real Truth about Low Graduation Rates, An Evidence-Based Commentary*. Urban Institute. https://www.urban.org/research/publication/real-truth-about-low-graduation-rates-evidence-based-commentary

[13] United States Census Bureau. (2022, July 1). *Quick Facts*. https://www.census.gov/quickfacts/VA

[14] U.S. Department of Education, National Center for Education Statistics. (2019). *Study of the Title I, Part A Grant Program Mathematical Formulas* (NCES 2019-016), Executive Summary, Introduction, and Chapter 1.

[15] Virginia Beach County Public Schools. (2023). *Free or Reduced Meal Prices.* https://www.vbschools.com/families/food-and-nutrition-services/free-or-reduced

[16] Virginia Department of Education. (2023). *Title I, Part A: Improving Basic Programs Operated by Local Educational Agencies.*https://www.doe.virginia.gov/programs-services/federal-programs/essa/title-i/title-i-part-a-improving-basic-programs-operated-by-local-educational-agencies

[17] Virginia Department of Education. (2023, November 13). *Diploma Graduates and Completers by School.* https://www.doe.virginia.gov/data-policy-funding/data-reports/statistics-reports/graduation-completion-dropout-postsecondary-data/high-school-graduates-completers/high-school-graduates-completers-archive-data

[18] Wilson, B., & Schneider, K.-J. (2023, January 24). *Improving admissions: how to evaluate the whole student.* Times Higher Education. https://www.timeshighereducation.com/campus/improving-admissions-how-evaluate-whole-student