2018

Identifying Areas for Change: A Case Study on North Carolina State Public School Performance

Olivia Leeson
Southern Methodist University, oleeson@smu.edu

Kelly Bean
Southern Methodist University, kbean@smu.edu

Jacob Drew
Southern Methodist University, jdrew@smu.edu

Follow this and additional works at: https://scholar.smu.edu/datasciencereview

Part of the Education Policy Commons

Identifying Areas for Change: A Case Study on North Carolina State Public School Performance

Kelly Bean[1], Olivia Leeson[1], Jacob Drew[1]

[1]Master of Science in Data Science, Southern Methodist University,
6425 Boaz Lane Dallas, TX 75205
{kbean, oleeson, jdrew}@smu.edu

Abstract.   In this paper, we present a framework to identify school-level factors within North Carolina public school administration's control that have a positive impact on school performance. Public school administrators struggle to improve the academic performance of their schools, as the most influential factors determining overall school performance are outside of their scope of influence. We consider the current circumstances responsible for poor performance in North Carolina public schools and their implications for future academic improvement. Our framework utilizes an extreme gradient boosting model to predict school performance scores using only school-level features that administrators can impact. By varying the inputs, administrators can estimate the potential improvements to school performance scores. We find that the number of short-term suspensions per 100 students in a school year is the most important feature used to estimate school performance scores, followed by the school's average daily attendance. Altering these features while holding all else constant is found to change school performance scores by just a few points. However, our framework creates an opportunity for schools to identify areas for change that may ultimately improve academic performance.

1   Introduction

Across the United States, public schools educate most of the nation's youth. Ninety percent of the roughly 56 million American students in grades 1-12 attended a public school during the Fall of 2017[1]. The nation's public schools, however, do not uniformly educate students to achieve the same levels of academic attainment. State averages vary significantly on standardized test scores for grades 4, 8, and 12 in reading, mathematics, science, and writing[2]. The same disparities are reflected within states: averages on standardized test scores and end-of-grade assessments vary significantly by school and school district[3].

---

[1] "Fast Facts: Back to School Statistics". National Center for Education Statistics. [Online.] https://nces.ed.gov/fastfacts/display.asp?id=372 [Accessed 1 March 2018]

[2] The Nation's Report Card. Grade 4 Mathematics 2017. [Online.] https://www.nationsreportcard.gov/profiles/stateprofile?chort=1&sub=MAT&sj=&sfj=NP&st=MN&year=2017R3 [Accessed 15 July 2018]

[3] "2016-17 Performance and Growth of North Carolina Public Schools: Executive Summary". North Carolina Department of Public Instruction. Accountability Services Division. 7

In North Carolina, school administrators have attempted to bridge divides in pupil academic achievement by targeting the lowest performing schools and creating plans to increase student performance on standardized tests and end-of-course grades. Despite investment in new initiatives, disparities in public school performance in the state remain prevalent [4]. The most important factors contributing to poor school performance are largely outside of school administrators' control, making improving school performance extremely difficult [1].

Current research indicates that poor academic performance is correlated with low household income and socio-economic status [2]. Insufficient funding for school resources has also been found to be correlated with poor school performance [3]. In North Carolina, federal, state, and Local Educational Agency (LEA) funding is mainly determined by the schools' number of registered students. However, schools generate additional funding through property taxes and donations, which are garnered at the local educational agency level. Schools in wealthier counties generate more dollars from taxable resources [5]. Schools in low income neighborhoods collect less funding than their medium-to-high income counterparts, and thus have less money to improve school resources and facilities [5].

Out of North Carolina's 2,531 public schools in the 2016-17 academic year, 505 schools qualified as low performing that year [3]. This is roughly 20% of all public schools. Some of those schools were among the 468 schools that also qualified as a recurring low performing school, making up 18.5% of schools in operation in the 2016-17 school year. A recurring low performing school must be identified as low performing in any two of the last three years of its operation [3]. As part of the state's effort to identify and improve public school performance, legislation was passed in 2013 that requires public schools to report certain school-level data. The data reported includes test scores, average daily attendance, and educator experience among other categories [3]. North Carolina now collects, for every public school, this data to assess the overall academic performance of each school in what is referred to as a School Report Card.

To identify and understand their relative impact on school performance, we present a framework that reveals the school-level factors contributing to overall school performance beyond the major factors of neighborhood socioeconomics and demographics. By discovering elements that are within administration's control, public schools can implement new policies to begin leveling the education gap that currently exists. The North Carolina Department of Public Instruction has already begun this initiative by mandating the collection of data [3]. Using the data from North Carolina Public School Report Cards and Statistical Profiles, we present a method for assessing

September 2017. [Online.] http://www.dpi.state.nc.us/docs/accountability/reporting/2017/documentation/exsumm17.pdf [Accessed 1 March 2018]

4 North Carolina Department of Public Instruction, District and School Transformation. [Online.] http://www.dpi.state.nc.us/schooltransformation/low-performing/2016-17/ [Accessed 1 March 2018]

5 "2018 Local School Finance Study". Public School Forum of North Carolina. [Online.] January 2018. https://www.ncforum.org/wp-content/uploads/2018/01/2018_PSF_LocalSchoolFinanceStudy-FINAL-PDF.pdf [Accessed 28 June 2018]

the school-level factors that can be changed through policy to improve school performance scores.

The first stage of the framework is to assess which school-level features should be included for analysis. The School Report Card is comprised of eight categories of information: school performance, test scores, school profile, educator experience and effectiveness, personnel experience and licensure, student demographics, school and classroom environment, and funding[6]. Educator experience tables, for example, contain information on the percentage of teachers with a certain number of years of experience. Environment tables include data on suspensions and crime ratios as well as data about wireless access points and student-to-internet-connected-computer ratios. Each category provides us with information that can potentially help explain differences in performance. However, not all categories are within school administrations' ability to change. Student demographics and funding cannot be reasonably altered by administration, and are thus purposefully removed from the framework. Test scores and information that is used to directly calculate school performance grade are also irrelevant to the framework.

The second stage of the framework utilizes a supervised regression model to predict school performance scores using inputs that can be impacted through policy changes. After the data processing we perform to organize the categories into one dataset, we are left with 90 features to use as inputs. The number of features is large, and the non-linear relationships between the features and school performance suggest that a nonparametric, nonlinear, multivariate regression is a reasonable approach to predicting school performance.

We use an extreme gradient boosting (XGBoost) algorithm for the model[7]. This algorithm makes predictions using an ensemble method that models the predicted errors of many decision trees to optimize final predictions[8]. Output of the model also reports the importance of each feature's influence in determining the final school performance score prediction[9]. The feature importance signals the impact - in absolute measures - each feature has on predicting school performance.

The XGBoost algorithm results in a model that explains about 65% of the variance in school performance score. The model can only explain 65% of the variance in school performance as we have purposefully removed inputs that also impact school performance but are irrelevant to the framework. These features include student demographics and funding [2][3] as well as the inputs that directly calculate school performance (standardized test scores, end-of-grade, and end-of-course scores[3]).

We find that the most influential factors in predicting the overall performance score include the school's average daily attendance percentage and the number of short-term suspensions per 100 students. The average daily attendance percentage has a positive

---

[6] North Carolina Department of Public Instruction. NC School Report Cards. [Online.] http://www.ncpublicschools.org/src/researchers/ [Accessed 10 July 2018]

[7] XGBoost Documentation. 2016. [Online.] https://xgboost.readthedocs.io/en/latest/index.html

[8] Introduction to Boosted Trees. Xgboost developers. 2016. [Online.] https://xgboost.readthedocs.io/en/latest/tutorials/model.html [Accessed 8 August 2018]

[9] Brownlee, Jason. Feature Importance and Feature Selection with XGBoost in Python. MachineLearningMastery. 31 Aug 2016. [Online.] https://machinelearningmastery.com/feature-importance-and-feature-selection-with-xgboost-in-python/ [Accessed 8 August 2018]

impact on school performance. The number of short-term suspensions per 100 students has a negative impact. Other factors of influence are the number of nationally board certified staff at the school (positive impact) and the percentage of teachers with 0-3 years of experience on the job (negative impact).

The relative impact of these features on school performance cannot be readily assessed by their feature importance from the XGBoost model. Feature importance measures are absolute, and do not signal whether relationships are positive or negative. To properly assess the influence of these features, the third stage of the framework involves predicting schools' performance given a simulated change in the input features' value. We simulate multiple scenarios under which these factors change, either singularly or jointly, and examine the predicted change in school performance grades. We find that school performance grades are improved by a few points by adjustments to average daily attendance and number of short-term suspensions per 100 students.

In Section 2 we present the current research into low performing schools and disparities in academic achievement. In Section 3 we explore how North Carolina state public schools are currently tackling the problem of poor school performance and the data the state is collecting to gain insights into its causes. In Section 4 we present the first step of the framework: data collection and feature selection. In Section 5 we discuss the second stage, using an XGBoost model to predict school performance scores irrespective of the schools' socioeconomics and demographics. In Section 5 we also report on the model's results and feature importance. The third stage of the framework: assessing feature impact, is discussed in Section 6. Section 7 reflects on the results of the framework. In Section 8, we discuss the ethical implications of our findings. Section 9 details our conclusions.

2   A Background on Factors Affecting Educational Attainment

2.1   School Funding

In the United States, public education funding is mainly the responsibility of state and local governments[10]. States have different systems of collecting and allocating funds to public schools. In North Carolina, as in most states, local property tax revenues are one of the most important sources of funding for school facilities[11]. Property tax revenue is dependent on a school district's economic infrastructure: wealthier counties can collect more revenue, which translates into more funds for schools in those counties [4]. Some research finds that school finance reform aimed at equalizing funding for all public schools helps close academic achievement gaps [4], while other research finds that there is not a strong relationship between school resources and performance [5].

---

[10]   "The Federal Role in Education". U.S. Department of Education. [Online.] https://www2.ed.gov/about/overview/fed/role.html [Accessed 17 August 2018]

[11]   "Basics of County Financing for Public Schools - NCACC." NCACC. [Online.] http://www.ncacc.org/DocumentCenter/View/2852. [Accessed 11 March 2018]

Research into the effects of school funding on school performance has been made possible by case studies of states that have attempted to neutralize disparities in school funding. Beginning in the 1970s, states including California, Arizona, Michigan, Kansas, Minnesota, New Jersey, and Texas adopted stances of fiscal neutrality in public education to make school funding more equitable. These states adopted the fiscal neutrality stance at different points after a 1971 California Supreme Court case, *Serrano v. Priest*. The court ruled that California's school finance system created excessive fiscal disparities between schools. The ruling called for an adoption of fiscal neutrality, which tasked the state to allocate public education funds equitably over school districts: "The quality of education may not be a function of wealth other than the wealth of the state as a whole" [6]. Several pieces of legislation that enforced these new fiscal approaches were later challenged in some states. In Texas, the case of *San Antonio Independent School District v. Rodriguez*, ruled in 1973 that the practice of fiscal neutrality was unlawful.

In New Jersey, *Robinson v. Cahill* (1973), followed by *Abbott v. Burke* (1985) led to some of the most distinct school finance reforms [7]. Since 1985, there have been ten iterations of the *Abbott v. Burke* case, each case attempting to push further toward a fiscal system of "equal educational opportunity" [7]. As a result of the initial *Abbott v. Burke* case, New Jersey was required to balance funding to schools across the state by increasing spending in the poorest districts. Later iterations included provisions for balancing preschool, summer school, and special needs programs [7].

The poorest school districts in need of reallocated funding in New Jersey have been coined "Abbott districts", and account for roughly 5% of all school districts in the state. Abbot districts educate nearly a fifth of New Jersey's student population and receive nearly 60% of the state's education funding[12]. The score gap between "Abbott" and "non-Abbott" districts has begun to close in recent years (from roughly 38 points in 1990 to 30 points in 2011)[13]. However, detractors claim that this relatively minor change in school scores is not enough to justify the expensive remedial orders [5].

The limited improvement of New Jersey's Abbott districts contributes to research that suggests that publics school funding is not the primary driver of educational achievement [5]. Across the nation there exist school districts that are high-poverty but high-performance, as well as those that are high-spending but low-performing [5]. This suggests that increasing public school funding alone is not a guaranteed way for states to substantially improve school performance.

2.2 Socioeconomic and Demographic Factors

School performance is strongly correlated with the socioeconomic standing of the school district [8]. Studies have shown that economically disadvantaged students on average do not perform as well academically as students who are not economically

---

[12] "Chris Christie claims 31 former Abbott districts receive 70 percent of the state aid," Politifact. [Online.] http://www.politifact.com/new-jersey/statements/2011/dec/01/chris-christie/Chris-Christie-claims-31-former-Abbott-districts-r/ [Accessed 11 March 2018]

[13] "NAEP New Jersey Assessment". National Assessment of Educational Progress (NAEP). [Online.] http://www.nj.gov/education/assessment/naep/nj.shtml [Accessed 11 March 2018]

disadvantaged [8]. Figure 1 depicts this relationship in North Carolina public schools for the 2016-17 school year. The trend shows that schools with increasing percentages of economically disadvantaged students tend to receive lower school performance grades.
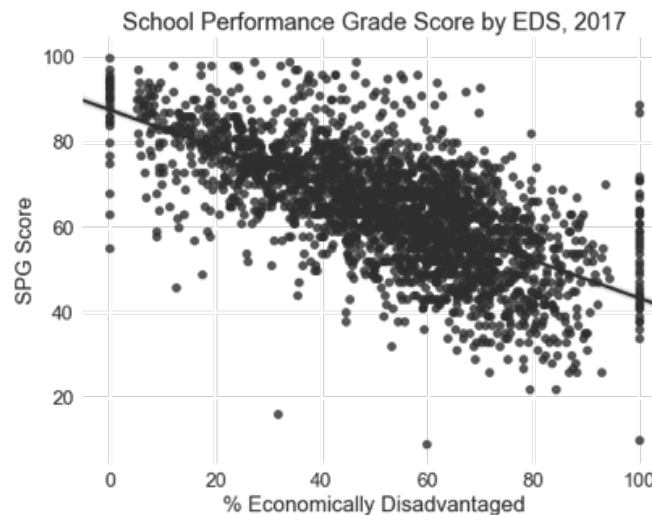

Figure 1. Scatterplot of school performance grade by percent of economically disadvantaged students. On average, as percent of economically disadvantaged students increase, School Performance Grade Score Decreases.

This link between poverty and low educational attainment has also been shown to be circuitous. Studies suggest that poorly educated populaces are more likely to receive lower incomes and live in low-income neighborhoods [8]. These low-income households are priced out of neighborhoods and school districts with more economic opportunity. The children of lower income households thus continue to receive a relatively poor education. A parent's education can influence the level of educational attainment a child receives. For example, it has been shown that a strong predictor of whether a student graduates from high school is whether their parents have earned college degrees [8].

Socioeconomic status has also been shown to be linked with race. In general, minority students are much more likely than whites to grow up in poverty [9]. The cyclical nature of education and poverty along with racially motivated housing policies perpetuate this link. Longstanding exclusionary housing policies have upheld an inertia to the poverty in certain areas. For example, in 1934, the Federal Housing Administration specifically prohibited its subsidized builders from selling homes to African Americans, and would refuse to insure mortgages in predominantly minority neighborhoods[14].

---

[14] Gross, Terry. "A 'Forgotten History' Of How The U.S. Government Segregated America." NPR. 3 May 2017. [Online.] www.npr.org/2017/05/03/526655831/a-forgotten-history-of-how-the-u-s-government-segregated-america [Accessed 11 March 2018]

When it comes to racial diversity in public schools, it is important to consider the common practice of segregation before the Civil Rights Movement. The city of Charlotte, North Carolina was the first to enact a district-wide bussing program after the 1971 court case, *Swann v. Charlotte-Mecklenburg Board of Education*[15]. The prime motivation behind bussing was not necessarily to improve school performance, but to increase racial integration in public schools. After decades of desegregation, a 2001 Fourth Circuity Court of Appeals ruling ended the mandatory bussing program, ordering districts to stop using race in pupil assignments[15].

The National Assessment of Educational Progress ran a study in 2015 to examine how racial segregation influences the black-white achievement gap. The report looked at the achievement of eighth grade students in relation to the percentage of black students in the school. The study found that the black-white achievement gap was largest in the highest density schools than the lowest[16]. This study suggests that there are positive impacts on education achievement in schools that are not highly segregated. However, given the current policies against bussing, the socioeconomic and racial demographic of a school's student body is outside of administration's control.

2.3   Attitudes Toward Education

The effects of individual, familial, and societal attitudes toward education have been studied in detail [8][9][10][11]. On the individual level, male students are more likely to drop out of high school than female students. Students with higher numbers of disciplinary incidents are also more likely to drop out of school [8]. Students measured as having greater self-control are expected to have fewer disciplinary events and a higher grade point average than those exhibiting limited self-control [10]. Female students with a high degree of self-control and fewer disciplinary incidents were measured to have the most optimistic view toward education and the positive benefits of attending college [10].

At the familial level, students with less-involved parents were more likely to drop out of high school than other students. These students were also more likely to drop out earlier (in the first two years of attendance) than other dropouts [11].

At the societal level, there exists a body of research which examines the so-called "oppositional culture" in minority and poverty-stricken peer groups [12]. This research shows that, in comparison to more affluent white and Asian students, black high school students typically spend less time on homework, are less likely to seek help when having trouble in school, and report lower perceived returns to education, along with lower educational expectations [12]. According to the same study, black students are less likely to report attending school because they enjoy classes, and typically spend less time on school activities or clubs [12]. More recent reports have shown evidence

---

[15]   Smith, Clint. "The Desegregation and Resegregation of Charlotte's Schools. The New Yorker. 3 October 2016. [Online.] https://www.newyorker.com/news/news-desk/the-desegregation-and-resegregation-of-charlottes-schools [Accessed 2 July 2018]

[16]   "School Composition and the Black-White Achievement Gap". National Center for Educational Progress. National Center for Education Statistics. 24 September 2015. [Online.] https://nces.ed.gov/nationsreportcard/subject/studies/pdf/school_composition_and_the_bw_achievement_gap_2015.pdf [Accessed 19 August 2018]

for some positive changes in these attitudes [12], but others point out that black students are still the most underrepresented ethnic category at American colleges, while Asians are the most overrepresented[17]. However, debates are ongoing as to whether these differences in student-body representations are due to the proposed "oppositional culture theory" or due to other systemic factors blocking groups of students from higher education. Even outside certain cultural underpinnings, other significant attitudes toward education at the societal level involve public impressions of minorities' ability to succeed [2].

Some research points out that teachers possess their own attitudes toward education, and more specifically, toward their students [12]. Minority students have been found to experience more suspensions, expulsions and other disciplinary events, and are otherwise viewed more critically than other students [2]. Even well-intentioned teachers can reinforce negative stereotypes; uncharacteristically positive feedback in response to a black student's good work or eloquent speech bolsters the idea that such performance is unexpected from minority students [2]. These attitudes at the individual, familial, and societal level are difficult to impact at the school level.

2.4   Interrelationships Among Factors

Each of the above sets of factors (school funding, socioeconomic, and attitudinal) are interrelated. Each pair of relationships contains a negative feedback loop which has served to expand educational inequality. For example, the relationship between school funding and socioeconomic factors is reinforcing. Poorer areas produce inadequate property taxes to help fund schools, and the poorly educated local workforce allegedly influences the educational deficiencies of local students [6]. These poorly educated students then become the poorly educated local workforce, and the cycle continues.

In turn, socioeconomic factors affect attitudes toward education, and vice versa. It has been hypothesized that the negative attitudes held by minorities stem from decades of explicitly racist policies that have discouraged minorities from actively participating in education [12]. If participation is viewed as being less likely to result in an equal and appropriate reward, then students are less likely to participate [12]. Students that participate less are less likely to graduate, and students that do not graduate are more likely to end up in low-income households [2]. These low-income non-graduates then produce deficient funding for local schools, and often help to reinforce negative sentiment toward education [12].

---

[17]   Ashkenas, Jeremy, et al. "Even With Affirmative Action, Blacks and Hispanics Are More Underrepresented at Top Colleges Than 35 Years Ago." The New York Times. 24 Aug. 2017. [Online]www.nytimes.com/interactive/2017/08/24/us/affirmative-action.html [Accessed 2 July 2018]

3 North Carolina School Performance

3.1 Measuring School Performance

The North Carolina Public School System uses a grading system to measure the overall performance of public schools. The calculation and reporting of school performance grade became a requirement in North Carolina since legislation on educational reform passed in 2013[3]. School Performance Grade (SPG) scores are defined by two factors: School Achievement Score and EVAAS (Education Value-Added Assessment System) Growth Score[3]. Final SPG Scores are a composite of the two weighted factors: School Achievement Score accounts for 80% of the final grade and EVAAS Growth score accounts for 20%.

School achievement score is calculated as the sum of points earned by a school on a series of achievement indicators. These indicators are end-of-grade assessments in reading, math, and science, along with end-of-course scores for math I and biology for intermediary schools. High schools are assessed on end-of-course math I, English II, and biology, along with ACT Scores, math course rigor, and 4-year graduation rate[3].

The EVAAS Growth Score is a measurement used to reflect how well a school improves test scores year over year. The EVAAS model used to calculate these scores is a product of the SAS Institute Inc. The system calculates a composite index of growth that is then used to determine the growth expectation and designation for each school. It does this by determining the achievement gap in end-of-course and end-of-grade assessments schools should meet year over year. That composite index is converted to a 100-point scale, resulting in the EVAAS growth score[3].

Table 1. School Performance Scores and Corresponding Grades.

Grade	Score
A	85-100
B	70-84
C	55-69
D	40-54
F	<40

The weighted combination of achievement and growth scores results in the SPG Score. This score is finally converted into a single letter grade, shown in Table 1, categorizing the performance of the school. Schools are designated as low performing if they receive an SPG of 'D' or 'F', and receive an EVAAS Growth Status of 'Met' or 'Not Met'. Schools can meet their EVAAS growth metric, but still be classified as low performing due to aggregate poor performance on the end-of-course and end-of-grade assessment.

3.2 Current State of North Carolina School Performance

In the 2016-17 academic year, roughly 20%, or 505, out of North Carolina's 2,531 public schools qualified as low performing. Some of those schools were among the 468

schools that also qualified as a recurring low performing school, making up 18.5% of schools in operation[3]. Out of the 2,617 public schools still operational in 2017, 902 of them had been classified as low performing for at least one school year in the previous four. That is rounded to 34% of schools. Meanwhile, 639 of those schools have been repeatedly low performing: 24.4% of all public schools have been low performing for at least two out of the four years between 2013 and 2017.

A recurring low performing school is defined as a school that has been classified as low performing at least twice in three years. The school does not need to be consecutively low performing within those three years to be classified as recurring low performing. For example, a school that is low performing in 2013-14 and 2015-16 but is not low performing in 2014-15 will be designated as a recurring low performing school in the 2015-16 school year. A low performing district is a district in which over 50% of schools are classified as low performing. Since reporting began in 2013, the 2016-17 school year is the first in which North Carolina has three consecutive years of SPG reporting. Table 2 shows that recurring low performing schools have been ticking up in the past three school years[3].

Table 2. Low Performing and Recurring Low Performing School Counts 2014-2017[3][18][19]

Designation	2014-15	2015-16	2016-17
Low-Performing School	481	489	505
Low-Performing District	15	10	11
Recurring Low-Performing School	401	415	468

School funding, demographics, and socioeconomics have been the most commonly explored and documented factors influencing school performance [1][2][3]. To quantify these influences on North Carolina Public schools, we explore the relationship between each and whether a school is low performing or repeatedly low performing.

3.2.1 Funding and School Performance in North Carolina

Data available for school-level funding in North Carolina is sparse. Public schools do not report their federal, state, or local per-pupil expense. The only school-level funding and expense data available in North Carolina are reported by the 168 charter schools that are required to report details of funding. Charter schools are not included in this framework, as they do not report the same data gathered by the school report card database.

---

[18] "2015-16 Performance and Growth of North Carolina Public Schools: Executive Summary". North Carolina Department of Public Instruction, Accountability Services Division. 1 September 2016. [Online.] http://www.dpi.state.nc.us/docs/accountability/reporting/2017/documentation/exsumm17.pdf [Accessed 1 March 2018].

[19] "2014-15 Performance and Growth of North Carolina Public Schools: Executive Summary". North Carolina Department of Public Instruction, Accountability Services Division. 2 September 2015. [Online.] http://www.dpi.state.nc.us/docs/accountability/reporting/2017/documentation/exsumm17.pdf [Accessed 1 March 2018].

The funding data reported for public schools are not the funds received by the individual school, but the average funds per-pupil received by the LEA, or Local Educational Agency, that the school belongs to. An LEA is synonymous with a school district. North Carolina has 115 LEAs: 89 of which operate in a single county. There are 11 counties in which multiple LEAs operate. LEAs receive funds through appropriations from county governments and private donations[20].

The mean per-pupil LEA funding for non-low performing schools is $2,221.17. The mean per-pupil LEA funding for low performing schools is $2,015.13. This is a difference of $206.04, which is not large. The reason for this is that both non-low performing and low performing schools operate in the same local educational agencies.
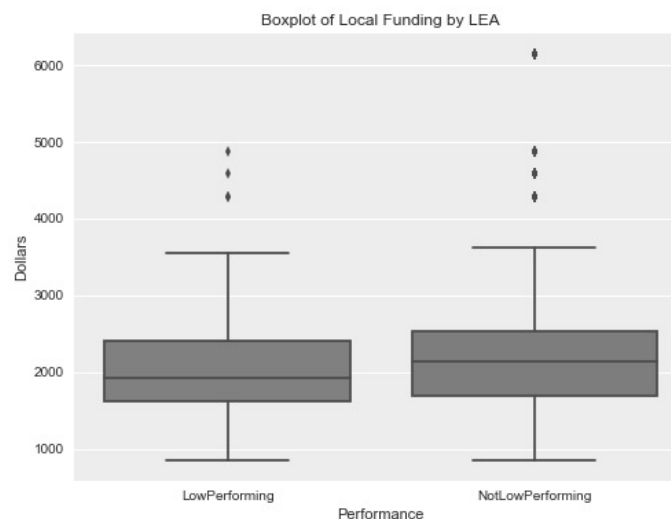

Figure 2. Boxplot distribution of per-pupil local funding at the LEA level for low performing and non-low performing schools

The difference in the mean per-pupil funding is attributed the Chapel Hill-Carrboro LEA. This LEA is represented by the outlier in the not low performing school group in Figure 2. Chapel Hill-Carrboro does not oversee any low performing schools, and the average per-pupil funding for this is LEA $6,150.80, significantly higher than the average LEA funding.

The lack of per-pupil funding data for individual schools means we are unable to quantify the effect of funding at the individual school level on school performance. From the data North Carolina reports, we can only see that there does not exist a large

---

20 Nordstrom, Kris. Financing Education in North Carolina A Budget and Tax Guide. NC Justice Center. 2017. [Online.] Available: http://www.ncjustice.org/sites/default/files/NCJC_education%20finance%20primer%20021 917.pdf. [Accessed 12 August 2018]

difference in LEA funding per-pupil between low performing and non-low performing schools.

3.2.2   Economically Disadvantaged Students and School Performance in North Carolina

North Carolina defines an economically disadvantaged (ED) student as a student who is eligible for free and reduced price meals under the National School Lunch Program. The eligibility criteria for this program is based on the household size and income. For a household size of one, lunch will be free for children whose guardian earns an annual income at or below $15,444. A household of size 4 is eligible for the program if annual income is at or below $31,590[21].

The distribution of the percentage of economically disadvantaged students varies between low performing, repeatedly low performing, and non-low performing schools, depicted in Figure 3.
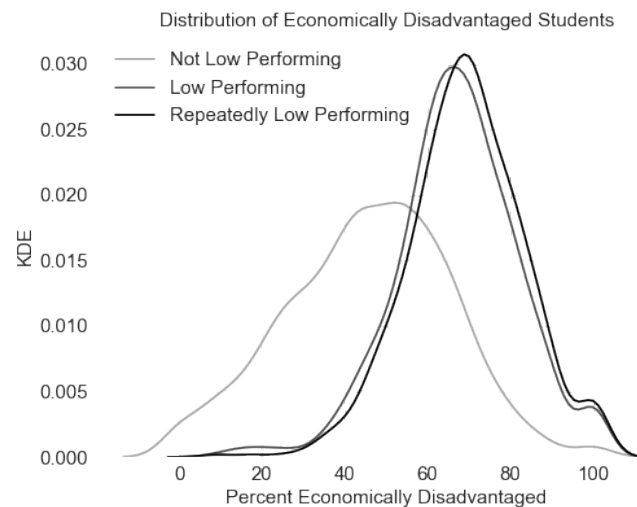

Figure 3. Density distribution plot of the percentage of economically disadvantaged students within schools either low performing, non-low performing, or repeatedly low performing.

The percentage of ED students in a school's student body is statistically different between low performing and non-low performing schools. A two-sided Welch's t-test between non-low performing and low performing school's percentage of economically disadvantaged students results in a p-value of less than .001. The t-test evaluates the distribution of the two groups along with their averages, and the p-value results tell us the probability of seeing the same differences in averages if we were to take a random

---

[21] Eligibility for Free or Reduced Priced Meals in the National School Lunch Program Announced. North Carolina Department of Public Instruction, Newsroom [Online.] http://www.ncpublicschools.org/newsroom/news/2016-17/20160729-01 [Accessed 14 May 2018]

sample of two averages from the same distribution. A p-value of less than .001 indicates a less than .001 probability that the average of the low performing school and non-low performing school's percentage of economically disadvantaged students is the same. We are more than 99% confident that on average, the percentage of ED students in low performing schools is higher than in non-low performing schools.

Table 3. Mean percent of economically disadvantaged students in low performing, non-low performing, and repeatedly low performing schools.

	Not Low Performing	Low Performing	Repeatedly Low Performing
% EDS	45.69	67.08	69.3

Table 3 summarizes the mean percentage of economically disadvantaged students in each type of school. The mean percentage of ED students in non-low performing schools is 45.69% while the mean percentage of ED students in low performing and repeatedly low performing schools is 67.08% and 69.3% respectively. On average, the difference in economically disadvantaged students between low performing and non-low performing schools is 21.39 percentage points.

3.2.3 Racial Demographics and School Performance in North Carolina

Each school in North Carolina reports the percentage of the total student body by racial backgrounds: white, black, Hispanic, Pacific Islander, Indian, Asian, and two or more races. The mean percentage of white students in low performing schools is 29.5%. In repeatedly low performing schools, white students represent 24.2% of the student body. Meanwhile, non-low performing schools are on average 58.6% white. In addition, non-low performing schools only have 19.1% black student body, while 40.6% of students represented in low performing schools are black and 44.9% of students in repeatedly low performing schools are black.
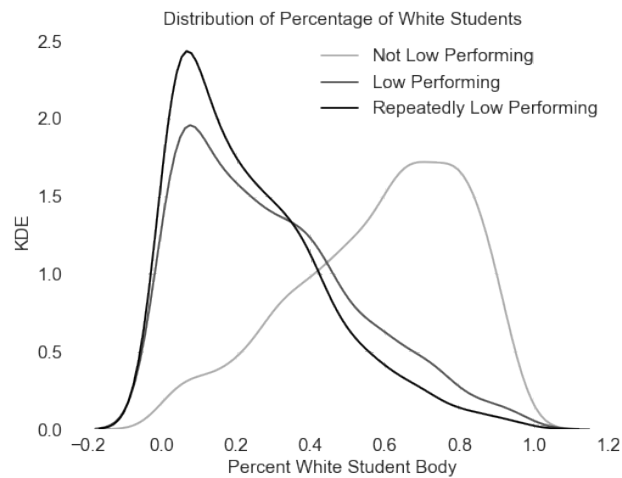
Figure 4. Density distribution plot of the percentage of white students within schools either low performing, non-low performing, or repeatedly low performing.

The two distribution plots of student body percentages of white and economically disadvantaged shows the difference between non-low performing and low performing schools. The graphs in Figures 3 and 4 suggest that any regression algorithm used for predicting low performing schools would be affected by the percentage of students by race and economically disadvantaged status.

4   Stage 1: Dataset Creation through Feature Selection

The data used in this case study are provided by the North Carolina Public Schools Report Card[6] and Statistical Profiles Databases[22]. The NC Public Schools Report Card is comprised of school and LEA level data collected for every school year since 2013. State legislation passed in that year required public schools to report on specific school-level metrics such as end-of-course and end-of-grade scores, graduation rates, and standardized test results. The data are categorized into separate tables, which we have compiled into one large dataset along with student demographic information from the Statistical Profiles Database for each school year between 2013-14 and 2016-17. All final datasets are processed to standards required for machine learning applications[23][24].

School Performance Grade Score (*SPG Score*) is our metric of interest for this framework. *SPG Score* is defined in Section 3, and is a composite score calculated by

---

[22]   Public Schools of North Carolina. Statistical Profiles. [Online.] apps.schools.nc.gov/statisticalprofile [Accessed 10 July 2018]

[23] Jacob, Drew. EducationDataNC. [Online.] https://github.com/jakemdrew/EducationDataNC

[24] Leeson, Olivia. Bean, Kelly. NCPDI-Capstone. [Online.] https://github.com/oleeson/NCPDI-Capstone

a schools' average scores on a variety of assessments. In the 2016-2017 school year, only 130 schools do not report an *SPG Score*. These are all magnet schools that are not required to adhere to the data reporting mandate, and are not considered in the framework along with charter schools.

The initial dataset for the 2016-2017 school year contains information on 385 school-level factors for 2,617 North Carolina public schools. The tables included are: School Profile, Profile Metrics, Funding, School Performance Grade, READY Accountability, Read to Achieve, Participation, School Indicators, Specialized Course Enrollment, College Enrollment, Environment, Personnel, Educator Experience, Educator Effectiveness, Statistical Profiles, Student Readiness, Economically Disadvantaged, Career and Technical Education, and Career and Education Credentials[25]. Appendix 1 provides a more detailed breakdown of each table and the data they contain.

We discard any features that are directly used in the calculation of *SPG Score*. Features discarded include those pertaining to specific test, end-of-course, or end-of-grade scores. Features related to EVAAS growth metrics are also removed.

Two final datasets are used for the *SPG Score* prediction and feature impact. The reduced feature set contains the Profile, Profile Metrics, Environment, Personnel, Educator Experience and Educator Effectiveness tables of the School Report Card. The reduced feature set is used in the model creation and validation for predicting *SPG score* in Stage 2. The complete feature set includes all the features of the reduced dataset, but also includes features from the Statistical Profiles, Economically Disadvantaged, and Funding tables. The complete feature set is used for feature impact assessment in Stage 3.

Tables of interest included in both datasets are the Environment and Educator Experience tables. The Environment table (Table 5) includes data on the average daily attendance, crime per 100 students, short and long term suspensions per 100 students, and ratio of students-to-internet-connected computer, among others. The Educator Experience table (Table 4) contains attributes for the percentage of teachers and principals at each school that have either 0-3 years of experience, 4-10 years of experience, or 10+ years of experience. The features belonging to these tables are of significant interest as they could be impacted by school policy changes.

Table 4. North Carolina Public School Report Card Educator Experience Table[25]

Attribute	Description
Year	School Year
Unit_code	School code
Experience	Values = 0-3 years, 4-10 years, 10+years
Pct_tch, LEA_pct_tch, st_pct_teach	Percentage of teachers at a given experience level at the school, LEA, and state level
Lea_pct_prin, st_pct_prin	Percentage of principals at a given experience level at the LEA and state level

---

[25] North Carolina Department of Public Instruction. NC School Report Cards Data Dictionary. [Online.] http://www.ncpublicschools.org/docs/src/researchers/data-dictionary.pdf [Accessed 10 July 2018]

Table 5. North Carolina Public School Report Card Environment Table[25]

Attribute	Description
Year	School Year
Unit_code	School code
Avg_daily_attend_pct	Average daily attendance percentage at school level *LEA and State level also available
Crime_per_c_num	Number of crimes or acts of violence per 100 students at school level *LEA and State level also available
Short_susp_per_c_num	Short term suspensions per 100 students at school level *LEA and State level also available
Long_susp_per_c_num	Long term suspensions per 100 students at school level *LEA and State level also available
Expelled_per_c_num	Expulsions per 100 students at school level *LEA and State level also available
Stud_internet_comp_num	Ratio of students to internet connected computer at school level *LEA and State level also available

We comprise separate datasets for each school year between 2013 and 2017. We find that year over year, North Carolina continues to collect more relevant data on school-level factors. The dataset for the 2016-17 school year is much more robust and includes more features than that collected for the 2013-14 school year. We use data from all four school years in the determination of feature importance in Stage 2. We use data only from 2016-17 in the model evaluation in Stage 2 and in assessing feature impact in Stage 3. The 2016-17 dataset is used for model evaluation in Stage 2 and feature impact in Stage 3 as this is the most robust and relevant school year we have access to data for. Our complete dataset for 2016-17 includes 2,313 public schools and 123 attributes. Our reduced dataset for 2017 includes 2,313 public schools and 90 attributes.

5   Stage 2: Modeling SPG Score and Important Features

5.1   The XGBoost Algorithm

XGBoost is a tree-based, gradient boosting algorithm. The purpose of the algorithm is to minimize a cost function relative to predicting a target variable. Tree-based methods function by splitting explanatory variables into bins which attempt to maximize "purity" (i.e. similarity) within each node, and a tree-shaped structure develops as nodes are recursively split up to a certain threshold[8]. The goal is to produce a decision tree which minimizes some measure of error between predictions and actual observed values. XGBoost expands upon this method by ensembling many decision trees which work together to produce a final prediction. An individual decision tree might determine an optimal split for each variable in a set of variables, but another decision tree could determine different splits.

For example, if we are to predict a school's performance grade score from per-pupil funding and average years of teacher experience, a decision tree could choose to first discretize per-pupil funding by splitting it into "high" and "low" nodes. Each of these two nodes are then split into "high" or "low" degrees of teacher experience in such a way that minimizes prediction error. The four final nodes are assigned a score corresponding to that node's most probable *SPG score*.

A second decision tree, however, could decide to first split average teacher experience, and would then split these nodes based on per-pupil funding. By switching the order of splitting the explanatory variables (or in some cases, by including different sets of explanatory variables), new trees can be produced. These new trees may even have similar predictive power, but they are also likely to produce different sets of predictions from one another. Tree ensembling works by allowing many decision trees to cast a "vote" in their predictions under the assumption that a majority of trees will produce accurate predictions even when others do not.

Other features within XGBoost take it beyond decision tree ensembles. A randomization parameter helps to reduce correlation among the trees, making trees' predictions more orthogonal, and the final predictions more accurate. Trees which are too complex to feasibly make predictions on unseen data are penalized, reducing overfitting. In minimizing the cost function (root mean squared error), XGBoost has parameters which allow each successive iteration to focus its efforts on those observations with the greatest prediction error.

This wealth of features means XGBoost typically performs better than traditional regression methods in terms of its predictive capability. However, the complexity of the model reduces interpretability[8]. Whereas a regression's predictors will produce a set of coefficients indicating their effect (both in magnitude and direction) on a target variable, combining many variations of decision trees to make a final prediction leads to a less immediately intelligible model.

However, a proxy for the straightforward coefficients of a regression model can be found in tree-based methods' feature importance metrics. When a node in a decision tree is split, we can calculate the following reduction in impurity, and attribute this reduction to the feature involved. When the tree is finished splitting nodes, those features with the largest proportional contribution toward decreasing impurity within nodes can be said to be the most "important." In other words, the factors that are most helpful in producing the most accurate predictions of *SPG Score* will be ranked most highly.

5.2 Predicting School Performance Grade Score with XGBoost

The XGBoost algorithm is used to model School Performance Grade Scores for all North Carolina public schools operating in the 2016-17 school year. As describe in Section 4, we have two feature sets to test the model's explanation of variance in *SPG Score* with and without funding and demographic features included. We must use two separate feature sets because our most important features are determined by the reduced feature set. Meanwhile, the assessment of those features' impact on SPG grade must be tested on a more realistic model that includes the funding and demographic features purposefully left out in the reduced feature set.
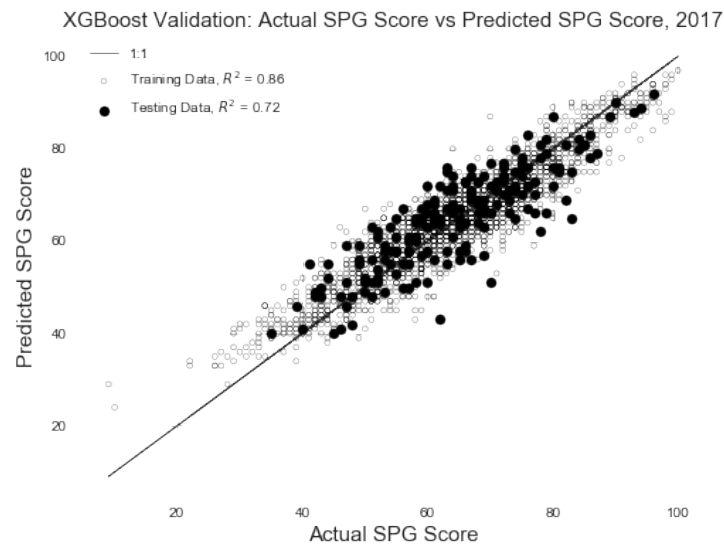
Figure 5. Performance of the XGBoost model using complete feature set; the closer each point is to the 1:1 line, the more accurate the model's predictions. The model achieves a mean absolute error of 4.98, indicating that out-of-sample predictions are accurate to within plus or minus 4.98 points.

The XGBoost model built from our complete feature set can accurately predict *SPG Score* to within 4.98 points on average. As shown in Figure 5, the model obtains an R^2 value of 0.72, indicating that 72% of the variance in *SPG Score* can be attributed to the list of input variables. The model built using our reduced subset of variables results in a mean absolute error of 5.5 and an R^2 value of 0.65.

The comparative results of the two models add evidence to our theory that funding and demographic features are good predictors for school performance. The complete feature set model can explain more of the variance in *SPG score* because funding and demographic features are included as inputs in prediction. The 65% of variance explained in the model using the reduced feature set still gives us confidence that the most important features resulting from the feature importance analysis are also good predictors of school performance. Meanwhile, the 72% of variance explained by the complete feature set gives us greater assurance that using XGBoost with these factors can generate realistic predictions of the expected changes to *SPG Score* when predicting from simulated datasets in Stage 3.

5.3 Determining the Most Important Features School Administrations Can Impact

Tree-based algorithms such as XGBoost can calculate the "importance" of each feature, indicating which features were most useful in producing accurate predictions. Within an individual decision tree, nodes are recursively split with a goal of minimizing prediction error. The importance of a given feature corresponds to the reduction in prediction error attained when splitting on the feature. If a certain variable is split multiple times, increasing the model's performance each time, it can be ranked as "more important" than a feature that only slightly improves the model when split in the decision tree. Because XGBoost ensembles many decision trees, feature importance is averaged across all the trees.

Table 6. Mean feature importance of top 8 features used to predict *SPG Score* on average between 2013-14 and 2016-17 school years using XGBoost algorithm.

Feature	Description	Feature Importance
short_susp_per_c_num	Short term suspensions per 100 students at school level	0.081
avg_daily_attend_pct	Average daily attendance percentage at school level	0.056
student_num	School size	0.045
lea_wap_num	Number of wireless access points at the LEA level	0.044
lea_avg_student_num	Average school size at the LEA level	0.041
category_cd_H	School educates high school grade levels only	0.032
tchyrs_0thru3_pct	Percentage of teachers with 0 to 3 years of experience at the school level	0.031
nbpts_num	Number of Nationally Board Certified Staff at the school level	0.030

The final feature importance calculated represents the average fraction of the total decrease in within-node impurity that each variable contributes. For example, a feature importance value of 0.05 indicates that the given variable accounts for 5% of the total decrease in the decision trees' impurity, averaged across all the decision trees used in XGBoost's ensemble. If all features were equally important in making predictions of *SPG Score*, given that nearly a hundred variables are included in the reduced set, we could expect values of 1%. Any feature importance value that surpasses this baseline proportion is reflective of a more highly significant variable.

Using the reduced set of features outlined in Section 4, we rank the factors with the greatest contribution toward producing accurate predictions of school performance. This feature set is applied to all school years between the 2013-14 and 2016-17. The mean feature importance is calculated across years to result in a ranking of most important features used to calculate *SPG Score*. The eight most significant factors are listed in Table 6: the school's number of suspensions per 100 students, average daily

attendance percentage, number of students within a school, the number of wireless internet access points in the school's district, the average number of students within a school across the school's entire district, whether the school contains high school students and classes only (no elementary or middle school attachment), the percentage of teachers with 0-3 years of teaching experience, and the number of nationally board-certified staff.

6 Stage 3: Assessing Impact of Important Features

Stage 3 of the framework involves assessing each important feature's impact on School Performance Grade Score. At this stage, we must use the complete feature set in modeling *SPG Score*. We iteratively modify important features by a range of multipliers to produce simulated datasets. XGBoost is used to predict *SPG Score* from these partially manufactured datasets, and the changes to *SPG Score* corresponding to each degree of multiplier are observed to deduce the underlying relationships.

For instance, each school's number of books per student is artificially increased by 10%, and, using the baseline model built on the original data, new predictions of schools' performance are made from the simulated data. The difference between the predicted *SPG Score* and the original *SPG Score* indicates whether a 10% increase in the number of books per student is expected to increase or decrease school performance, and by what magnitude. This process repeats with, for example, a 5% increase, a 0% change, and then decreases of -5% and -10%, each time calculating the predicted change in performance. Another factor is chosen, and the iterative simulated changes and subsequent predictions repeat. The factors are then ranked in order of those which require the smallest changes in exchange for the largest increases in *SPG Score*.

As an extension, some factors which are seen to have their own underlying interrelationships (for example, schools' number of suspensions and average daily attendance percentages) are jointly altered through a range of multipliers. While it is useful to examine the fundamental relationships between school performance and explanatory factors individually, it is also advantageous to view how some factors affect *SPG Score* in the context of their interactions with other factors.

Factors are ranked by those with the greatest increases in *SPG Score* given a correspondingly small change in the factor. Note that as depicted in Figure 6, some factors see increases to *SPG Score* when they are increased (average daily attendance percentage) while other factors see performance gains when decreased (number of suspensions).
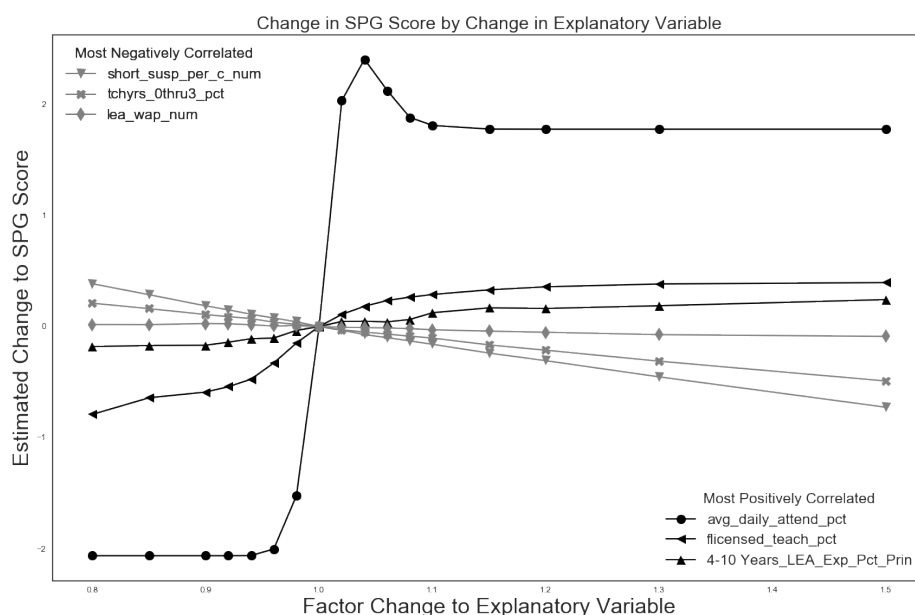
Figure 6. Predicted improvement in *SPG Score* (y-axis) at each degree of simulated changes to original data (x-axis). The "most negatively correlated" factors are the three which, when decreased, correspond to the largest increases in *SPG Score*. The "most positively correlated" are those which, when increased, see the largest performance gains.

Two sets of variables among the factors identified as having a high feature importance are jointly iterated through a similar range of multipliers. The average daily attendance percentage and the number of suspensions per 100 students are the two opposing factors with the greatest impact on *SPG Score* predictions. Together, these features possess an obvious tradeoff: attendance must necessarily decrease if suspensions are to increase. Similarly, the percentage of teachers with 0-3 years of experience, and the percentage of teachers with 4-10 years of experience will contain a certain degree of compromise. Viewing these sets of variables in unison elucidates how they cooperatively affect school performance.
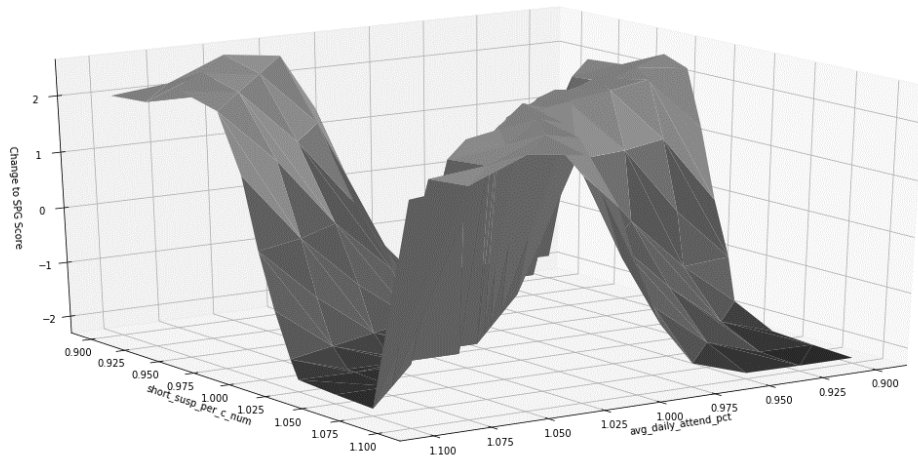
Figure 7. Predicted improvement in *SPG Score* (vertical axis) at each degree of simulated changes to the number of suspensions (left horizontal axis) and average daily attendance (right horizontal axis).
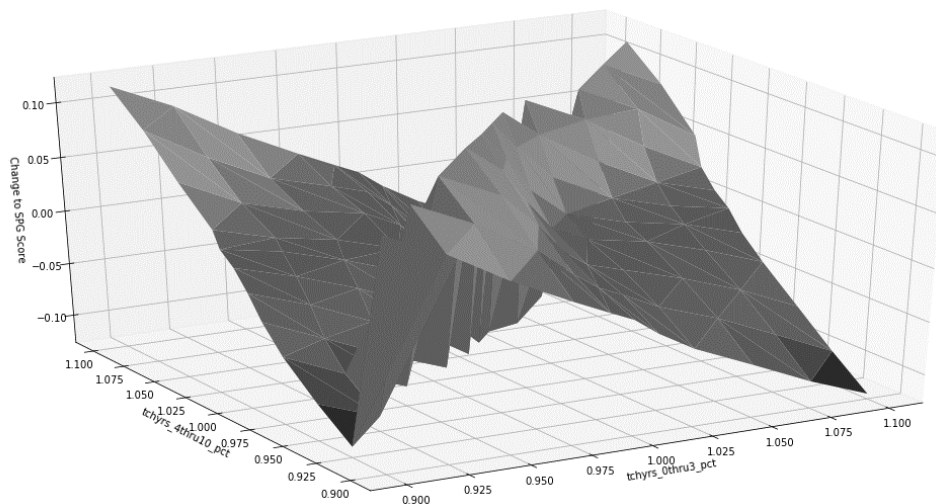

Figure 8. Predicted improvement in *SPG Score* (vertical axis) at each degree of simulated changes to the percentage of teachers with 4-10 years of experience (left horizontal axis) and the percentage of teachers with 0-3 years of experience (right horizontal axis).

7 Analysis of Results from Framework

7.1 XGBoost Prediction Validation Analysis

As seen in Figure 4, the model using our complete feature set can explain 72% of the variance in *SPG Score*. A mean absolute error of 4.98 indicates that our predictions on out-of-sample data typically fall within plus or minus 5 points of the actual *SPG Score*. The training set R^2 of 0.86 suggests that 86% of the variance in *SPG Score* can be explained by our model. In other words, the data our model initially uses to develop a basis for future predictions has a high degree of accuracy. The test set of data withheld from the initial model development phase confirms this with an R^2 of 0.72. Removing the demographic- and funding-related variables led to a test set R^2 of 0.65 and mean absolute error of 5.5 points.

Although perfect predictions are more ideal, there are some degrees of natural randomness that cannot be accounted for in the data. For our purposes of deriving feature importance, and of ranking features by their impact on *SPG Score*, forecasting the target variable to within five points is highly adequate. This high degree of accuracy gives us greater assurance that using XGBoost with these specific factors can generate realistic predictions of the most important features, and of the expected changes to *SPG Score* when predicting from simulated datasets.

7.2 Feature Importance Analysis

Table 6 indicates that the two most important factors in accurately predicting *SPG Score* are related to student attendance. The number of suspensions per 100 students, and average daily attendance percentage, collectively account for 15% of the total features' importance. This intuitively makes sense: regardless of any positive effect that other factors could have, and independent of any benefit a school and its administration can offer in terms of educational capability, if students are not present to take advantage of them, performance will suffer.

An interpretation for the two next most related variables (number of students, and district average number of students per school) is less obvious. Appendices 2 and 3 depict the scatterplots of correlation between school size and school performance. Further analysis shows a slight a positive relationship between student population and school performance, but that the variance in *SPG Score* is much greater in schools with fewer students.

Several points of speculation arise from these relationships: it could be the case that, within our model, the number of students is serving as a proxy for some other, more direct relationship among factors that we have intentionally excluded, thus circumventing our attempt to reduce the appearance of non-manipulable factors. *SPG Score* has a clear negative relationship with the percentage of students that are economically disadvantaged (*EDS*). Appendix 4 includes the scatterplots of the relationships between *EDS*, *SPG Score*, and other important features. Schools with more economically disadvantaged children perform worse. The relationship between the size of the student population and the percentage of those which are *EDS* shows that

larger schools tend to have a smaller portion of students in poverty. These transitive correlations make it possible that the ability of factors related to school size to predict *SPG Score* stems from their corresponding reflection of some other underlying relationships (seen here as the percentage of students which are economically disadvantaged).

A similar effect is seen in the relationship between the number of wireless access points (another important feature) and the percentage of economically disadvantaged students; schools with a poorer student body have fewer access points. In fact, each of the "important variables" in predicting *SPG Score* show relationships with the percentage of students which are economically disadvantaged: poorer schools have higher numbers of suspensions, lower attendance, fewer students, fewer wireless access points, a greater percentage of teachers with 0-3 years of experience, and fewer teachers which are nationally board-certified. Scatterplots of relationships can be found in Appendix 5. The only important factor with a somewhat neutral relationship with *EDS* is *category_cd_H*, identifying whether a school contains high school classes and students only (otherwise including middle and elementary school classes).

Although some of these variables show definite relationships with the proportion of economically disadvantaged students, the statistical weakness of these relationships detracts from the validity of this proxy theory, that each of the features outlined above are simply reflecting the immovable factors that we intentionally exclude from analysis. Except for the direct relationship between *SPG Score* and *EDS*, which obtains an R^2 of 0.47, the average R^2 attained by the remaining variables with *EDS* is only 0.098.

7.3   Assessing Important Features Analysis

Regardless of the possible interpretations for the important features, the *SPG Score* predictions at each degree of simulated data (Figure 6) exhibit the relative magnitudes of each variable's effect on school performance. Although each of the factors in the dataset were iteratively altered and predicted from, including the important features, not all the important features ranked highly enough in their *SPG* impact to be listed.

The variable with the greatest predicted *SPG Score* improvement when increased is the average daily attendance percentage. Daily attendance is far and away the most impactful, and similarly shows the greatest detriment to school performance when decreased. Given that it has a fairly weak relationship with the confounding *EDS* factor, and given that the entirety of the remaining variables have been held constant in observing these changes, the effects of attendance on *SPG Score* are likely independent of a proxy effect.

Similarly, the number of suspensions is ranked the highest among the variables which, when decreased, correspond to the greatest improvement in *SPG Score*. The joint predictions in Figure 7 confirm that the most significant increases occur when attendance is maximized and suspensions are minimized, but additionally indicates that there exists a wider range of "acceptable" numbers of suspensions when attendance is sufficiently increased. The opposite situation however, is not true—if attendance is not sufficiently high, then proportionally larger decreases to suspensions are needed to mediate the decreases to *SPG Score*.

The next two variables with the greatest impact on school performance are the percentage of nationally licensed teachers, and the number of teachers in the district with 4-10 years of experience. Additionally, the percentage of teachers with 0-3 years of experience is the second most impactful variable when decreased, meaning that teacher related variables account for half of the top increases to *SPG Score* given a relatively small change. Evidently, increasing the number of more experienced teachers, or decreasing the number of less experienced teachers, corresponds to increased school performance. The joint predictions seen in Figure 8 reinforce this but contribute additional nuance. The greatest gains are seen when the percentage of experienced teachers are increased and inexperienced teachers are decreased (and vice versa). However, there exists a "ridge" mediating both, indicating that schools which can effectively diversify their teaching staff, even without massive recruiting or layoffs of teachers with different levels of experience, can still improve their performance.

Whether these important features are simply reflecting other underlying factors (such as poverty) cannot truly be known, but the relationships between them and *EDS* highlight the inseparability of many school-related features. If it is the case that these features are exclusively serving as proxies for other underlying factors, then they reinforce the sturdiness and depth of such hidden elements, as well as the far-reaching implications of their interwoven nature. The most likely explanation is that, for most of these important factors, there is a certain feedback among themselves, and between them and other more integral factors; just like the primary factors discussed in Section 2. Suspensions affect attendance, which affects student performance; poorly performing schools have a harder time attracting experienced and qualified teachers, which further affects performance.

8   Ethics

Education is arguably the most important factor that contributes to an individuals' future career, family life, and wellbeing. North Carolina's Public School's vision is that "every public-school student will graduate ready for post-secondary education and work, prepared to be a globally engaged and productive citizen"[26]. This inherited responsibility is important, as most factors that can affect a child's education happen outside of school jurisdiction. It is thus the role of the state and local government to make decisions within their purview to ensure that students of all races, creeds, and backgrounds receive an adequate education.

Ethical issues could arise from the implementation of new policy aimed to increase school performance based on our findings. Those issues include fairness and avoidance of discrimination. The ACM code of ethics lists 'Be Fair and Take Action not to Discriminate' as a general ethical principle of computing professionals[27]. Education is a different domain than computing, but the same principle exists.

---

[26] North Carolina Department of Public Instruction, Mission and Requirements. [Online.] http://www.ncpublicschools.org/organization/mission/ [Accessed 1 March 2018]

[27] ACM Code of Ethics and Professional Conduct. [Online.] https://www.acm.org/code-of-ethics [Accessed 19 August 2018]

Our results suggest that decreasing the number of short term suspensions per 100 students can have positive returns to school performance grade. Decreasing the number of short term suspensions would require a change in policy and discipline. The line for which behavior constitutes a short-term suspension would need to be redefined.

Currently, schools use suspensions as a tool to attempt to eliminate behavior that distracts other students from learning. The consequences of a more lenient policy may detract from other well-behaved students from getting the best education possible. Keeping students with poor, distracting behavior in class to lower short-term suspensions at the detriment of other well-behaved students could be viewed as unfair. The consequences of which may also end up lowering the academic performance and attainment of other students in class.

Given the risks associated with keeping poorly behaved students in the classroom, an alternative policy change may end up discriminating against certain students. For example, if administrators plan to both reduce short-term suspensions and avoid distracting well-behaved students from learning, they may need to segregate the poorly behaved students from the rest of the class. This policy would likely discriminate against young black men, as this population of students are more likely to be suspended from school[28].

Similarly, increasing average daily attendance would require a policy change on behalf of the school. Currently students may not be able to attend class because they are ill, must take care of siblings, or are truant due to frustration with the school system, among other reasons. School administrators may want to tighten attendance requirements for passing students to the next grade or graduation. The students who are not able to meet these attendance requirements are likely the students from economically disadvantaged households. The outcome of the policy would discriminate against these groups of students. Additionally, the method for increasing attendance would require additional policy that would not necessarily solve the underlying problems causing the absences in the first place.

Changes in policy may lead to discrimination and possible issues of fairness and equity amongst student populations. It is very important that any policy changes resultant from this framework constantly re-evaluated. A change to one feature in the system can cause unexpected changes in relationships amongst other features. Lowering the number of short-term suspensions may result in an increase in long-term suspensions, for example. It is paramount that the results of policy changes are also reevaluated through the framework.

9   Conclusions

This framework provides evidence toward the importance of data reporting in public schools. The insights drawn from modeling school performance could not be achieved

---

[28] Hui, Keung T. "Are black students more likely to be suspended in your child's NC school district?". News Observer. 15 January 2018. [Online.] https://www.newsobserver.com/news/local/education/article194706934.html [Accessed 20 August 2018]

without using school-level factors that can be adjusted under policy as inputs. Without the data to begin with, we would not be able to separate school-level factors from the demographics of the student body. Separating school-level factors allows school administrators to identify the symptoms of the underlying causes responsible for school performance. These symptoms can then be applied to larger concepts important to improving school performance; for instance, keeping students in seats. Those greater concepts unrelated to race or socioeconomics are more readily supported with the data under this framework.

The framework we've developed provides insight into school-level features that impact school performance. Based on the connections associated with attendance and teacher experience discovered, administrators can better understand the underlying factors that contribute to school performance. It is important to note that given our analysis, it is likely that the features we've found to be important are still related to the more deeply-rooted drivers of academic performance discussed in Sections 2 and 3 of this paper. The daily attendance percentage, number of short-term suspensions per 100 students, and teacher experience still reflect a school's ability to keep students in seats, deal with behavioral issues, and pay salaries for experienced teachers. It may not be the case that directly altering any of these variables will solve the underlying issues. However, in a situation with so many interrelated components, paying attention to some of these uncovered symptoms could lead the way toward approaching the more fundamental causes.
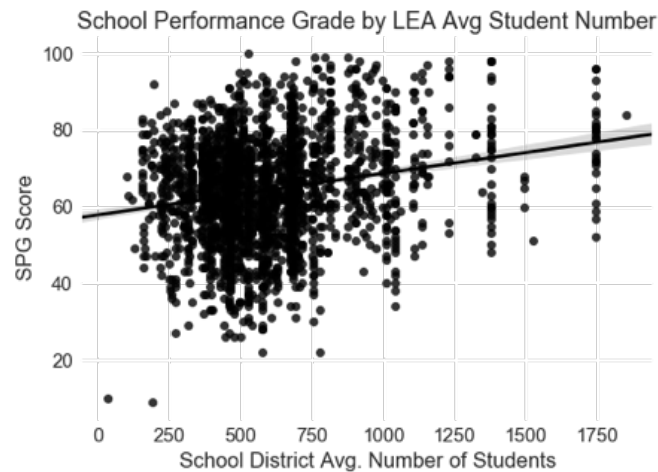
Appendix

Appendix 1. Data tables from North Carolina Public School Report Card and Statistical Profile datasets[6].

Table	Num Factors	Description
Profile	35	Address, Calendar Type, ESEA status, Category, Student Number
Profile Metrics	18	LEA name, Category, Size of class for each grade
Funding	39	Total Expense, Percent of Expense spent on School, LEA, and State levels
School Performance Grade	24	SPG for all reporting elements, EVAAS Growth score, Graduation Rate
READY Accountability	884	EOG, EOC, ACT, Graduation Rate broken down by 12 student demographics for all schools
Read to Achieve	10	Number of students by testing outcome at School, LEA, and State level
Participation	10	Participation targets assigned and met
School Indicators	27	SAT, IB, AP, participation targets at School, LEA, and State level
Specialized Course Enrollment	17	AP, IB, and CTE courses offered at School, LEA, and State level

| College Enrollment | 12 | Percent and count of students enrolled on college out of graduating class |
Environment	43	Daily attendance, short and long term suspensions, number of crimes, internet access
Personnel	72	Principal and Teacher demographics including race, sex, education, licensure, quality
Educator Experience	3	Percentage of teachers and principals at a given experience level
Educator Effectiveness	66	Level and standard of principals and teachers scored on 1-8
Statistical Profiles	25	Racial and gender demographics per school
Student Readiness	2	Percent of student body proficient at grades 6 and 9
Economically Disadvantaged	5	Percent of students economically disadvantaged
Career and Technical Education	15	Courses offered in career and technical education
Career and Technical Education Credentials	5	Performance/credentials offered for technical education courses

Appendix 2. Regression plot of School Performance Grade Score by the average number of school size in the Local Education Agency.

Appendix 3. Regression plot of School Performance Grade Score by school size.
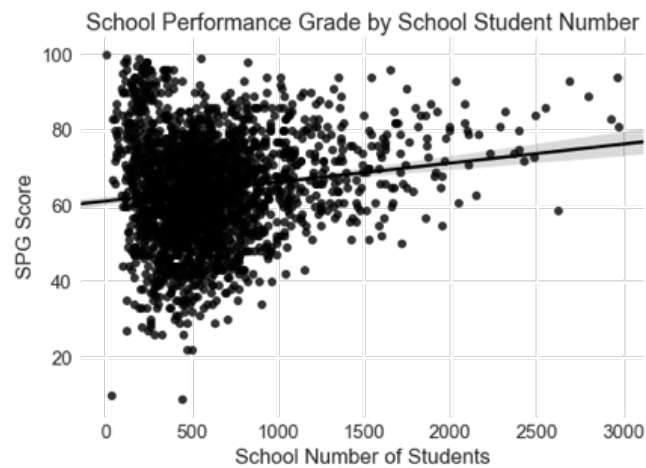
Appendix 4. Top left: regression plot demonstrating the strong negative correlation between *SPG Score* and the percentage of students that are economically disadvantaged. Each of the following plots demonstrates strong correlations between the "most important" factors.
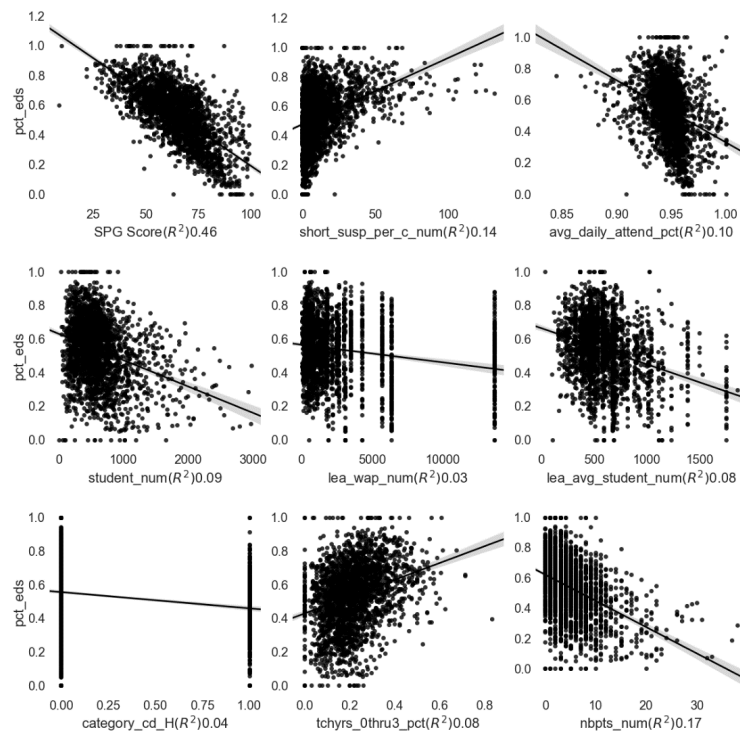
Appendix 5. Bottom right: Demonstrates relationship between schools' percentage of expenditures on salaries and the percentage of students economically disadvantaged (poorer schools spend proportionally less on salaries). Each other plot shows the relationship between an expenditure-related category, and *SPG Score* (only salary-related expenditures correspond to an increase).
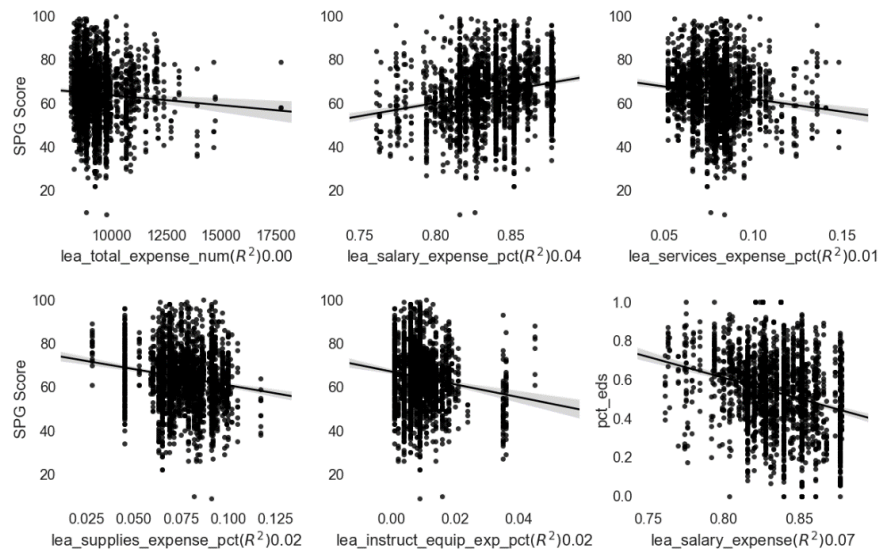
References

1. Okpala, C.O., Smith, F., Jones, E., Ellis, R.: A clear link between school and teacher characteristics, student demographics, and student achievement. Education 120(3), 487 (2000)
2. Merolla, D.M.: Oppostitional culture, black habitus and education: A new perspective on racial differences in student attitudes and beliefs. 21(1/2), 99-111 (2014)
3. Crampton, F.E., Thompson, D.C., Vesely, R.S.: The forgotten side of school finance equity: The role of infrastructure funding in student success. NASSP Bulletin 88(640), 29-52 (2004)
4. Card, D., Payne, A.A.: School finance reform, the distribution of school spending, and the distribution of student test scores. Journal of Public Economics 83(1), 49-82 (2002)
5. Hanushek, E.A.: Courting Failure: How School Finance Lawsuits Exploit Judges' Good Intentions and Harm Our Children. Stanford: Education Next Books (2006)
6. Tractenberg, P.L.: New Jersey: Robinson v. Cahill: The "Thorough and efficient" clause. Law and Contemporary Problems 38(3), 312-332 (1974)
7. Howard, J.E.: Abbott v. Burke: An historical analysis of school finance reform in New Jersey. Ph.D. thesis (2006).
8. Morris, T.O., West, E.: Individual Family, and School Factors Indicative of Caring: Predicting High School Graduation for NLTS2 Students with ED. Ph.D. thesis (2017)
9. Orfield, G., Lee, C.: Why segregation matters: Poverty and educational inequality. UCLA: The Civil Rights Project (2005)
10. Tangney, J. Baumeister, R., Luzio Boone, A.: High self-control predicts good adjustment, less pathology, better grades, and interpersonal success 72, 271-324 (2004)

11. Bowman-Perrott, L., Burke, M., de Marin, S., Zhang, N., Davis, H.: A meta-analysis of single-case research on behavior contracts 39 (09 2004)
12. Harris, A.L.: I (don't) hate school: revisiting oppositional culture theory of blacks' resistance to schooling 85, 797-833 (12 2006)