Project Goal
Predict 2017-18 TEA accountability scores

Summary
The following project includes a written preliminary analysis of the relationship between TEA accountability scores and variables like poverty, STAAR test scores, and attendance. This analysis served as a foundation for predicting 2017-18 TEA accountability scores using random forest and multiple regression machine learning models.

Included
Original research paper and additional predictive model research

Texas Public Education Accountability

Group project
I completed the paper, research, and compiled data set
Additional group members completed visualizations

Abstract - The goal of the project was to examine the correlation between Texas Education Agency school district accountability scores and poverty levels. Texas Education Agency, TEA, officials have publicly stated the relationship between the two is weak while providing no documented research or analysis. Our research found evidence of a moderate to large correlation between district accountability scores and poverty levels. We examined other variables relationship to district accountability scores, such as funding and expenditures, attendance, and STAAR standardized test results. Our analysis was limited to examining distributions and linear relationships or correlations between two variables. To gain more insight, can we build predictive models to understand and identify the most important variables linked to accountability and standardized test scores? While limited to a single year of data due to recent changes in TEA scoring methods, we hope this research will drive future analysis as additional data is available.

I. Introduction

What is an education accountability system and what is the purpose? "… Accountability systems can serve many purposes, including sharing information, measuring progress toward state and local goals and supporting greater educational equity." Ultimately, each state accountability system is intended as a foundation to measure school quality. Education accountability programs are composed of standards of achievement, measurement of student performance against these standards, and consequences for not meeting the standards (50-State Comparison: States' School Accountability Systems).

Legislation has shaped and continues to shape education accountability standards. The legislative history spans back to the 1965 Elementary and Secondary School Act that was reauthorized in 2002 as the No Child Left Behind Act and again in December 2015 as the Every Student Succeeds Act, or ESSA. The ESSA grants states more power over the state accountability standards than in the past. As a result, states are taking the opportunity to better describe school quality, beyond test results and graduation rates (50-State Comparison: States' School Accountability Systems).

In Texas, the Texas Education Agency, TEA, is responsible for K-12 public education. The TEA oversees school district operations and its work is shaped by laws created by the Texas Legislature and U.S. Congress. The TEA is managed by the State Board of Education and its commissioner is appointed by the governor. Every two years the Texas Legislature passes a biennium budget. Often, along with new budget approvals, as the composition of the legislature, public perception, and research change, so do the Texas accountability standards (About TEA).

The purpose of this project was to explore the relationship between poverty and the current Texas A-F accountability system for school districts during the 2017-18 school year. Additionally, we examined other variables to determine their relationship to accountability scores. We examined different sources of funding and how they are being spent, along with attendance, and STAAR standardized test results. Standardized tests are commonly criticized across the country for not measuring student knowledge but rather the characteristics of the school and district communities. These criticisms of standardized tests fuel criticisms of state accountability systems because these systems often rely heavily on standardized test scores.

II. Problem

"TEA commissioner Mike Morath said the new grading method produces scores that 'are not strongly correlated with poverty,' a common criticism of A-through-F grading systems in the 13 other states that use them (Webb)." In our project, we seek to understand the relationship between the percent of economically disadvantaged students, attendance rates, funding, and STAAR results.

First, we need to better understand the accountability system. The current Texas A-F accountability system was created with 2017 legislation passed by the Texas Legislature. Within the system, both K-12 campuses and districts receive a raw score and an A-F letter grade. The scores are composed of two major categories: 70% is student achievement or school progress, whichever is higher, and the remaining 30% of the rating is closing the gaps. Additionally, given the multi-layer and conditional structure of the TEA accountability score formula, it is unclear how much STAAR results contribute to district scores or individual campus scores.

Student achievement or school progress make up the first major category at 70% of the overall accountability score. For elementary and middle school campuses, student achievement is based on STAAR test results. For high schools and districts, student achievement is based on STAAR test results, graduation rates, and college, career, and military readiness. Graduation rate is 20% of the total score and STAAR and readiness are each 40% of the total student achievement score.

If higher, school progress scores are used instead of student achievement scores. School progress includes academic growth and relative performance. Academic growth compares students performance on the STAAR mathematics and reading exams year over year. Relative performance measures campus or district performance against similar campuses or districts.

The remaining 30% of the accountability score is closing the gaps. Closing the gaps is a measure of the difference between the highest and lowest performing demographic groups. The metrics within close the gaps are not solely based on STAAR scores and can also include graduation rates, attendance, discipline, and English language proficiency.

In addition to the TEA standards and accountability scores, or measurement of students against these standards, the accountability system includes consequences. If a school or district receives F scores for two consecutive years they must submit an improvement plan to the TEA. After receiving F scores for three consecutive years schools face closure and/or districts are required to appoint a board of managers to govern the district (2018 Accountability Manual).

Ultimately, the data downloaded from the TEA was chosen based on the components of the accountability score detailed above. Much of the data is from data sources like the Texas Academic Performance Report, TAPR, and the Public Education Information Management System, PEIMS reports. Additional financial data, including tax rates and revenue, was also included to gain context. For example, how much does a district spend on instruction as a percent of the overall actual expenses?

Also, in order to discern if poverty and accountability scores are correlated, we need to define poverty. In education, economically disadvantaged is the measure of poverty used for campuses and districts. Economically disadvantaged students are those who are eligible for free or reduced-price meals under the National School Lunch and Child Nutrition Program. The program identifies eligible students based on annual household income. In 2018, for example, if a household of one earns less than $22,459, or a household of eight earns less than $78,403 a year, the students in the household are considered economically disadvantaged. The National School Lunch guidelines are updated yearly and differ from the Federal Poverty guidelines. The Federal Poverty guidelines for households of one and eight are $12,140 and $42,380 respectively (Child Nutrition Programs: Income Eligibility Guidelines).

III. Data

The TEA data analyzed for the project includes: A-F raw and letter accountability scores, STAAR, or State of Texas Assessments of Academic Readiness standardized test results, enrollment, demographics, percent economically disadvantaged, attendance, drop out & graduation rates, teacher & administrator data, dyslexia, tax revenue, expenditures, and maintenance and operation costs. The data analyzed is for a single school year, during the most recent A-F accountability system. While the A-F accountability system was implemented during the 2016-17 school year, the state changed the formula and format of the accountability scores for the 2017-18 year, likely due to the ESSA. Once the 2018-19 school year data is released in the Fall of 2019, there will be two years of comparable data available for analysis.

All project data is publicly available through the Texas Education Agency, https://tea.texas.gov/. The primary data are accountability and STAAR district scores for the 2017-18 school year. Overall, there are approximately 1,200 districts, 1,000 public and 200 charter school districts.

Accountability scores: School district number, name and type (independent school district or charter), region number and name, county name, enrollment, economically disadvantaged percent, LEP percent (Limited English Proficient), accountability raw score and letter grade (overall, student achievement, school progress, academic growth, relative performance, closing the gaps), and distinction post secondary readiness

STAAR scores: test results for all subjects all grades, all grades English language arts, and all grades mathematics

Additional TEA data was joined with the accountability and STAAR scores, resulting in 1,013 observations and 99 variables total. The files are listed below and are ordered by appearance in the SAS EG file. Prior to loading and joining the additional data, redundant columns, like district enrollment, were deleted. Additionally, for all Texas Academic Performance Report, or TAPR data, the TAPR Data Dictionary was used to decode the column names (2017-18 TAPR Download District).

Enrollment: student enrollment by demographic groups, economically disadvantaged
Attendance: rates of attendance by demographic groups
Staff: teacher, administrator, salary, years experience, number of staff
Tax Revenue: federal, state, local district tax revenue, cost per student
Functions: categories of expenditures like teacher salary, transportation, food

service, administration, and extracurricular activities
Tax Rates: funds for maintenance, operations, and payment on debt which finances the district facilities, including maintenance and operations tax rate and interest and sinking tax rates
Graduation: graduation rates by demographic groups (all, African American, Hispanic, white), additional demographic groups are not included because all groups are not represented and therefore not included in the TEA accountability scores
Dyslexia: percent of dyslexic students in district

Both independent school districts, ISD, and charter schools are represented in the accountability and STAAR data. However, charter schools are not included after joining all of our data because they cannot levy taxes, instead they receive solely public funding from the state. Additionally, including charter schools in our overall analysis is problematic because of low enrollment and the lack of representation in various demographic and socioeconomic groups.

IV. Data cleaning/Validation

TEA data is highly regulated, subject to data governance standards and audits. Therefore, data cleaning was not necessary, however, both a major hurricane and student privacy legislation impacted our data.

In our initial data exploration we found missing data and missing accountability scores for Houston area school districts. During the 2017-18 school year, Hurricane Harvey hit Houston on August 25-29, 2017. The TEA identified 109 school districts, directly affected by Hurricane Harvey, as eligible to be labeled as 'Not Rated' for the 2017-18 school year (2018 Hurricane Harvey Provision). However, whether or not a school district received a 'Not Rated' for their letter grade, it still received a raw score. If a raw score was available for these schools, we included them in the analysis.

Additionally, data is masked within the TAPR data in order to conceal performance results where the sample size is too small and could lead to student identification. Based on the masking rules for the TAPR data, if the values of any TAPR variable are between 0 and 5 the values are changed in the data to -1 to comply with the Family Educational Rights and Privacy Act. In our data the -1 values were replaced with 2 values, the approximate average of 1, 2, 3, and 4 values (Explanation of 2017–18 TAPR Masking Rules).

Decisions regarding what variables to keep or discard were also made. Due to the large stand deviation present in the number of enrolled students per district, using the raw count of students in various categories of data made comparisons between districts impossible. While scaling the data was an option, we felt that measurements in units of standard deviations lacked the interpretability of percentages. Ultimately, we dropped variables with raw values and retained those with percentages. This allowed for clear and interpretable visualizations.

V. Analysis

For our data analysis, visualizations were completed in SAS, as required by the assignment, and data collection, filtering, and joining were completed in R.

In our initial data analysis, we found low enrollment school districts have a lack of representation among economic and demographic groups. Variables with high rates of missing values among demographic groups included: 4 year graduation rate for African Americans and Attendance for Pacific Islander. Both of these demographic variables are used in calculating student achievement and closing the gaps, two major categories in calculating school district accountability. To remedy this, we filtered the original data set. The resulting data set included school districts with over 1,000 students and had 510 observations from the original 1,018.

Next, we examined variable distributions and found the majority of the continuous variables were normally distributed. Figure 1 shows the distribution of the letter grade accountability scores, with a right skew. Additionally, box plots were used to compare distributions of different segments. Figure 2, the box plot examining the distribution of economically disadvantaged students by accountability score segments, shows accountability scores decreasing as the percent of economically disadvantaged students in the district increase.

We continued our analysis with the additional variables using scatter plots and Pearson correlation. Scatter plots were used because the majority of variables of interest are continuous. Correlations noted are Pearson correlations, measuring the linear relationships, and raw accountability scores were used in the following plots (Figures 3-9).

First, percent economically disadvantaged to accountability scores and are moderately to largely negatively correlated ($r = -0.528$) and the STAAR score and accountability score are strongly to largely positively correlated ($r = 0.8264$). The additional variables relationship to poverty and accountability scores were also explored. Attendance rates and the percent economically disadvantaged are moderately to largely negatively correlated, $r = -0.551$, while attendance rates and accountability scores are moderately to largely positively correlated ($r = 0.5713$). For expenditures, the cost per student and accountability scores have no linear relationship (r = -0.017). Both average teacher salary and percent of budget designated for instruction have small to weak positive correlations to accountability scores (r = 0.2026, r = 0.3508). Instruction includes "...all activities dealing directly with the interaction between teachers and students, including instruction aided with computers... and expenditures to provide resources for Juvenile Justice Alternative Education Programs... (2010–11 AEIS Glossary)."

VI. Visualizations

See appendix for visualizations.

VIII. Suggestions

Overall, additional research is required in order to better understand and improve public education accountability systems. There are various types of accountability systems in the United States. Are some state accountability systems more successful in conveying and measuring school quality? In our limited research, we examined distributions and linear relationships or correlations between two variables. To gain more insight, can we build predictive models to understand and identify the most important variables linked to accountability scores?

Future studies:

Which types of state accountability systems are better at conveying school quality? Types of accountability systems include: A-F, descriptive (Needs Improvement, Average, Good, Great, Excellent), index (1-100, 1-10, or 1-5), tier-of-support (Comprehensive Support and Improvement, Targeted Support and Improvement, None)

Accountability and Reporting: Current System

Can we predict district TEA A-F accountability scores?

Can we predict district TEA A-F accountability scores without STAAR test results?

Can we predict STAAR test scores?

How is dyslexia district data related to STAAR English language arts exam test results?

Are Hispanic graduation rates important in predicting TEA A-F accountability score?

Are expenditures by extracurriculars significant in predicting TEA A-F accountability scores?

IX. Conclusion

In addition to the consequences the TEA imposes based on accountability scores, school districts across the country use performance on standardized tests and accountability scores to make decisions. Decisions that include student promotion to the next grade level, student eligibility to participate in advanced coursework, eligibility to graduate high school, and teacher tenure. Our research shows there are moderate to large correlations between accountability scores, STAAR results, and the percent of economically disadvantaged students. This result coincides with research from around the country that found test scores can be predicted "by looking at some of the important characteristics of the community, rather than factors related to the schools themselves, like student-teacher ratios or teacher quality (Students' test scores tell us more about the community they live in than what they know)."

There are also limitations in the TEA data. Some of the most important questions about children's family life and other factors outside the classroom are not included in the data. For example, does a child have to work part-time during high school in order to support themselves or their family? Are a child's parents or guardians involved in their education? Do they read to their children? What are the parents or guardians educational background? Do the children have access to transportation in order to get to extracurricular activities?

Ultimately, school districts need to be partners in future research. Some Texas school districts recently took the opportunity to be a part of a TEA pilot program, an alternative to the legislated A-F accountability system. As a part of the program school districts define their own accountability systems and submit them for review. However, this program places a high burden on school districts with potentially no reward. In order to fully implement the proposed accountability system, the TEA must approve the district designed systems for use beginning in the 2018-19 school year (20 school systems are part of local academic accountability system pilot).

If a Texas school district constructs an interesting accountability system, it would be difficult to conduct additional research. For the majority of school districts and campuses, the TEA prescribes the data, format, and frequency at which districts submit data that contribute to accountability scores. Therefore, even if a school district started to measure additional variables to better describe district and school quality, the TEA would need to require the data to obtain comparable data across Texas.

Additional research is also necessary to understand the costs and benefits of the other types of accountability systems currently used in the United States. In addition to the overall effectiveness, one essential component of these accountability systems are the ways in which the system ratings are conveyed. For example, there is a stigma associated with districts receiving a C letter grade. Students, parents, and teachers assume a C is barely passing. The previous accountability scores provided by the state of Texas included the following categories: exemplary, recognized, acceptable, unacceptable. Is receiving an acceptable rating more palatable than a C rating?

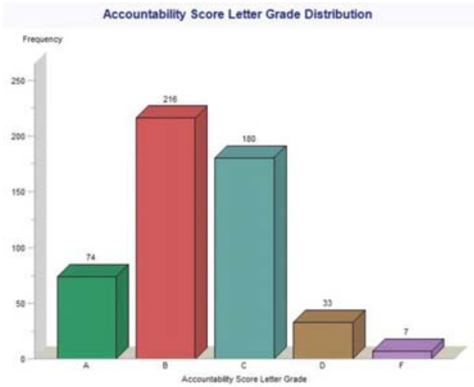


Figure 1

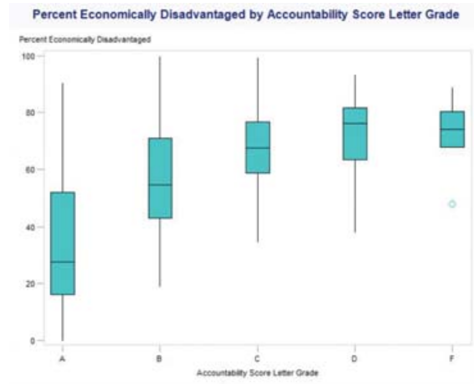


Figure 2

Figure 3

Figure 4

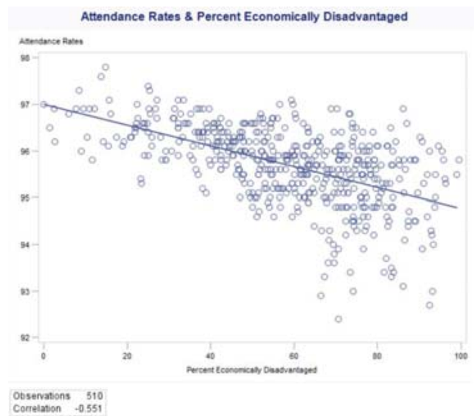


Figure 5

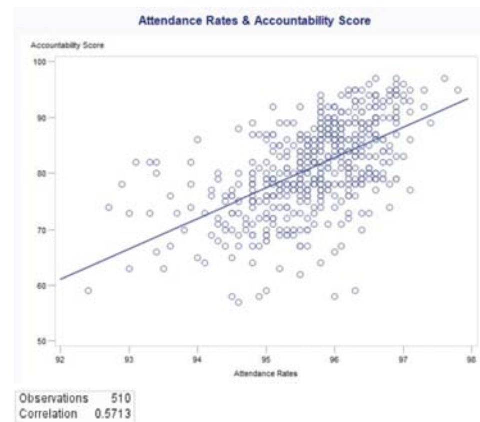
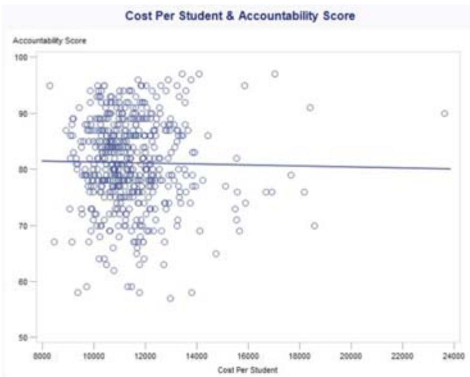


Figure 6

Accountability and economically disadvantaged
1: Accountability score letter Grade distribution (bar plot)
2: Percent economically disadvantaged by accountability score letter grade (box plot)
3: Percent economically disadvantaged and accountability score (scatter plot)
STAAR exam
4: STAAR score and accountability score (scatter plot)

Attendance
5: Attendance rates and percent economically disadvantaged (scatter plot)
6: Attendance rates and accountability scores (scatter plot)
Expenditures
7: Cost per student and accountability scores (scatter plot)
8: Average teacher salary and accountability scores (scatter plot)
9: Percent of budget designated for instruction and accountability score (scatter plot)

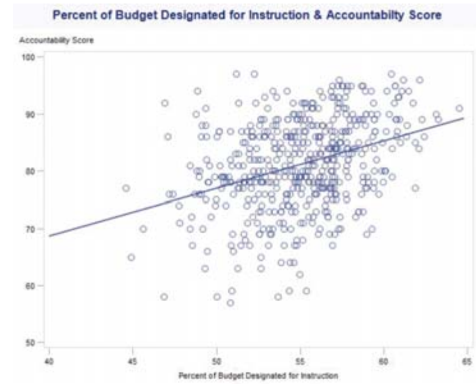


Observations 510
Correlation -0.017

Figure 7

Observations 510
Correlation 0.2026

Figure 8



Observations 510
Correlation 0.3508

Figure 9

Additional research

In the initial machine learning models, the more complex random forest model performed better than the less complex, multiple regression model. The higher the MAE, mean square error, MSE, mean square error, and RMSE, root mean square error, the worse the model. Errors for the two models are listed below.

Random Forest
MAE: 5.24 points
MSE: 58.02
RMSE: 7.62 points

Multiple Regression
MAE: 11.42 points
MSE: 202.06
RMSE: 14.21 points

Variable importance from the random forest model:
1. Instructional expenditure ratio
2. Percent economically disadvantaged of enrolled
3. Enrollment in 9th grade percent
4. Attendance all percent
5. Enrollment Disciplinary Alternative Education Program percent
6. All school leadership function percent
7. All percent local tax
8. Teacher turnover percent
9. Enrolled mobile percent
10. Campus administrative salary percent