Introduction

Why are these education data sets interesting?

Preliminary analysis on the 2005 and 2014 Texas school districts' performance on standardized tests revealed some interesting findings. Students passed the Texas Assessment of Knowledge and Skills (TAKS) in 2005 at an average rate of 86.88 percent. In 2014, Texas students passed a new state standardized test, the State of Texas Assessment of Academic Readiness (STAAR), at an average rate of 76.29 percent. There is a significant difference between the two percentages, and this piqued our interest.

Many factors can affect school performance, and we sought to explore how different ones--teacher experience, school funding, and demographics, for instance--affected the percentage of Texas students passing the standardized test in each district. We hoped that this would shed some light on how and why the performance of students dropped between 2005 and 2014. Perhaps the 2014 STAAR was simply harder than the 2005 TAKS, or perhaps it favored certain educational inputs over others. In the end, we aimed to provide some rigorous analysis on the types of inputs and policy that should be implemented in order to maximize student performance.

TAKS Versus STAAR

The Texas Department of Education changed the state-mandated standardized test in 2012.[1] The new STAAR includes more hours of testing and a higher level of difficulty in order to let exceptional students shine.[2] The new test has been criticized because it has strict time limits, counts towards final class grades in high school, and places an uneven burden on students who already struggle--minority and low-income students.[3] Our analysis shed some light on these criticisms and helped us understand the differences between the tests.

---

[1] Jeffrey Weiss, "STAAR vs. TAKS: Texas' new standardized tests come to schools next week," The Dallas Morning News (Dallas, TX), March 19, 2012, News.

[2] "A Comparison of Assessment Attributes Texas Assessment of Knowledge and Skills (TAKS) to State of Texas Assessment of Academic Readiness (STAAR)," chart, Texas Education Agency, [page 1].

[3] "Top Ten Problems With the STAAR Test," Save Texas Schools.

Literature Review

What factors affect school performance?

If we were to determine what factors lead to higher school performance, we first had to understand some of the variables that are thought to affect learning. Our literature review provides some research and theory about the variables that may contribute to school success.

Stiefel, Schwartz, and Ellen found that differences in standardized test scores can be attributed to poverty, school size, and attendance.[4] The Economic Policy Institute found in a review of available research that there is overwhelming evidence that schools with high rates of economically disadvantaged students perform worse on standardized tests.[5] The institute identified a variety of causal pathways that contribute to this relationship.[6] Race is also an indicator of school achievement. Stanford University's Center for Economic Policy Analysis reports that the 2012 achievement gap between white students and black and hispanic students was a significant 0.5 and 0.9 standard deviations.[7] Demographics and socioeconomic distributions obviously matter in school performance, and our model accounted for that.

Mackenzie's research also shows that the amount and type of funding a school receives affects its performance.[8] He found that more funding improves school achievement and that a higher percentage of local funding is also correlated with increased achievement.[9] He theorized that local funding meant greater local control over education, and thus education better suited to the particular students in the school district, eventually leading to higher-achieving students.[10]

---

[4] Leanna Stiefel, Amy Ellen Schwartz, and Ingrid Gould Ellen, "Disentangling the Racial Test Score Gap: Probing the Evidence in a Large Urban School District," Journal of Policy Analysis and Management 26 (2006).

[5] Richard Rothstein, The Racial Achievement Gap, Segregated Schools, and Segregated Neighborhoods – A Constitutional Insult(District of Columbia, DC: Economic Policy Institute, 2014).

[6] Rothstein, The Racial Achievement Gap.

[7] Racial and Ethnic Achievement.

[8] John Mackenzie, Public School Funding and Report (Newark, DE: University of Delaware, 2006), [Page 1].

[9] Mackenzie, Public School Funding and Report, [Page 7].

[10] Mackenzie, Public School Funding and Report, [Page 1].

MacKenzie's research conscientiously ignored SAT scores and opted for another level of student performance: the previously mentioned National Assessment of Educational Progress. He showed that varying participation rates in different states made comparing SAT scores an unreliable method for studying student performance.[11]

Additionally, in a 2010 report the Urban Institute's Center for Analysis of Longitudinal Data in Education Research reviewed studies showing the connection between teacher experience and student performance.[12] They found evidence that teacher experience matters more than degree level and other qualifications, but that it matters most in the first few years of teaching.[13]

The last study we reviewed outlines a model similar to the one we will be using. The Relationship of School Inputs to Public School Performance in New York State by Herbert J. Kiesling examines many different variables affecting school performance on tests. His report analyzed 17 variables; five of them are directly related to this report: teacher-pupil ratio, median teacher salary, school district size in average yearly attendance, school property value per pupil, and median years of teacher experience in the school district.

When Kiesling used several different regression models, he found varying results about the significance of school inputs on student performance. He suggests that the study should have looked at students' performance over a period of time, rather than a single year. Kiesling concluded that more work needs to be done in order to find the perfect formula of independent variables associated with students' performance on standardized tests.[14]

---

[11] Mackenzie, Public School Funding and Report, [Page 3].
[12] Jennifer King Rice, The Impact of Teacher Experience Examining the Evidence nd Policy Implications, issue brief no. 11 (District of Columbia, DC: CALDER, 2010).
[13] Rice, The Impact of Teacher, [Page 1].
[14] Herbert J. Kiesling, The Relationship of School Inputs to Public School Performance in New York State., technical report no. P-4211 (n.p.: ERIC, 1969).

Our research seeks to build on Kiesling's model. In doing so, we hope to gain some clarity on what contributes to students passing standardized tests. Our model is certainly not fully comprehensive, but it is well-researched and we hope it incorporates some of the highest-stakes factors in considering school achievement.

Hypotheses

Drawing from previous works and our analysis, this paper will examine the following hypotheses:

1. Higher teacher's incentive (average teacher's salary) positively influences school district performance (Kiesling, 1969).
2. Higher total revenue per pupil positively influences school district performance (Kiesling, 1969; Mackenzie, 2006). *[Note that we found total revenue per pupil and total district-owned property per pupil to have high correlation (>.90), therefore we chose to include only total revenue per pupil, whose data is more complete, in this hypothesis.]*
3. Higher student-teacher ratio (average class size) negatively influences school district performance (Kiesling, 1969).
4. Higher attendance positively influences school district performance (Kielsing, 1969).
5. Larger school district size negatively influences school district performance (Driscoll, 2006).
6. Greater human capital (average teacher's experience) positively influences school district performance (Kiesling, 1969).
7. Higher percentage of local funding positively influences school district performance (Mackenzie, 2006).

Data Analysis

For our analysis we looked at two different sets of data on Texas school districts from different points in time. One dataset is from 2005 and another is from 2014. Having two models increases our overall knowledge about the school inputs that most affect student achievement. Moreover, the two datasets will help us better understand how implementation of policy--namely the switch from TAKS to STAAR--may have changed school inputs, student achievement, and the connection between the two.

2005 Data

We used the 2005 District and Chartered Detailed Education dataset, which contains data from 1,044 public school districts in Texas. The dataset includes variables related to student performance, district finances, and teacher and staff characteristics. We selectively chose a set of variables to construct our models in order to test the above-mentioned hypotheses. Table 1 shows the descriptive statistics for each variable, and Table 2 shows the Pearson pairwise correlation matrix.

[Refer to Table 1.1 “2005 Descriptive Statistics”]

[Refer to Table 2.1 “2005 Pearson’s Pairwise Correlation Matrix”]

For dependent variables, our analysis opted for percentage of students passing state mandated academic skills tests (passall) and average SAT score (satscor). The state mandated tests, or Texas Assessment of Knowledge and Skills (TAKS), are a set of standardized tests to assess students’ attainment of reading, writing, math, science, and social studies skills required under Texas education standards. According to Texas Education Agency, a standardized score of 2,100 out of 3,000 (may vary for each test) is required to pass the exams.[15] On average, 86.88% of students in Texas school districts passed TAKS in 2005.

15 “A Comparison of Assessment,” chart.

SAT score is a national standardized test based on mathematics, critical reading, and writing. It has a full score of 1,600. The average score of Texas school districts was 968.69 in 2005, which was a little below the national average of 1,050.[16] According to Mackenzie's studies, even though SAT scores are more widely used, they are not alone suitable to reflect students' performance since not all students take the test. Thus in this paper, we use SAT score as a complement to the percentage of students' passing standardized tests, and not as the major criterion of our studies.

Our independent variables included average teacher salary, revenue per pupil, class size, average daily attendance, district size, average teacher experience in number of years, and percent of revenue from local sources. District size was measured in total number of students rather than number of schools in order to prevent difference in school sizes from affecting the estimation. Mean teacher salary across Texas school districts was $36,570 in 2005, mean revenue per pupil was $7,907, mean class size was 12 students, mean daily attendance rate was 96.13%, mean district size was 3,941 students, mean teacher experience was 12 years, and mean percent funding from local sources was 45.86%.

Revenue per pupil, district size, and percent revenue from local sources all had great amount of variation. Revenue per pupil ranged from $1,762 to $53,775; district size ranged from 20 to 210,670 students; and percent funding from local sources ranged from 0% to 99%.

Demographic and socioeconomic variables, namely percentage of African American students, Hispanic students, and economically disadvantaged students, were also added as controls. Note that Hispanic students and economically disadvantaged students control variables have a strong, statistically significant correlation (0.67). This could mean an overlapping of the two demographics and thus a possibility for multicollinearity.

---

[16] PowerScore Test Preparation, The Old SAT Vs. The New SAT Test.

2014 Data

In order to further our analysis, we analyzed 2014 data from the same organization. Our 2014 model looked at 777 school districts in Texas. Table 1 shows the descriptive statistics for each variable, and Table 2 shows the Pearson pairwise correlation matrix.

[Refer to Table 1.2 “2014 Descriptive Statistics”]
[Refer to Table 2.2 “2014 Pearson’s Pairwise Correlation Matrix”]

Our dependent variables were the same for the 2005 data as the 2014 data: percent of students passing standardized tests and average SAT score. Between 2005 and 2014, though, the standardized tests and the SATs changed. As noted before, the state-mandated tests in Texas were changed from the TAKS to STAAR. In addition, a writing section was added to the SATs and it was scored out of 2400 points instead of 1600 points.

For parity, our analysis included the same independent variables for the 2014 data as it did for the 2005 data. Mean teacher salary across Texas school districts was $48,938.80 in 2014, mean revenue per pupil was $10,869, mean class size was about 14 students, mean daily attendance rate was 95.69%, mean district size was 6,327 students, mean teacher experience was still 12 years, and mean percent funding from local sources was 43.65%.

The large range in school district size in this dataset mirrors the 2005 data; the smallest school district was 103 students while the largest was 210,716 students. Percent funding from local sources a range almost identical to the 2005 data with a minimum of 0% and a maximum of 98%. Lastly, revenue per pupil had a large range similar to the 2005 data with a minimum of $6,553 and a maximum of $64,422.

Demographic data provided shows that on average 8.5% of students in each school district were African American, 40.8% were Hispanic and 57.8% were economically disadvantaged. (Refer to table 1.2)

It should be noted that some of the independent variables used in our model have significantly high standard deviations. For example, the percent of African American students, Hispanic students, and economically disadvantaged students all have extremely high standard deviations. This suggests that some schools have a larger population of different races compared to other schools. (Refer to table 1.2)

Before and After

Comparing descriptive statistics from 2005 (Table 1.1 and 1.2) and 2014 (Table 2.1 and 2.2), one can see that average total revenue per capita increased by $2,961, average teacher salary increased by $9,368 and average total number of student per district increased by $2,385. When we account for the change in SATs from a 1600-point format to a 2400-point format, we see that SAT score dropped by 16 points. It is also noteworthy that percentage of revenue from local sources decreased by 2.20% and attendance fell by 3.97%.

One of the most notable differences in the descriptive statistics for the 2005 and 2014 data is in the increased average teacher salary. From 2005 5o 2014, teacher salary increased roughly $10,000. There are numerous factors that could lead to such a drastic increase. One factor to be considered is inflation. Although with the economic recession and recent drops in inflation rates, it is doubtful that the $10,000 increase can be attributed solely to this variable. Another possibility is that there was an increased amount of funding towards teachers salaries. This could mean that there was an increase in the percentage of funding going towards salaries from the years 2005 to 2014. While the minimum teacher salary increased about $5,000, the maximum increased by about $16,000. This high maximum--perhaps caused by a few very highly-paid teachers--may be to blame for the marked increase in mean average teacher salary.

Research Model

First of all, one must establish the baseline models for the relationship between school inputs and performance. Kiesling suggested a multiple regression model based on theoretical importance and statistical consideration (factor analysis was conducted to remove variables that tend to cause multicollinearity problem).[17] Other demographic variables such as race and socio-economic status are also added as controls.[18] We adopted a modified version of this model as our baseline:

$$[1a]\ (passall)_i = \beta_0 + \beta_1(salteach)_i + \beta_2(revpup)_i + \beta_3(class)_i + \beta_4(pafr)_i + \beta_5(phisp)_i + \beta_6(pecd)_i + \varepsilon_i$$

$$[1b]\ (satscor)_i = \beta_0 + \beta_1(salteach)_i + \beta_2(revpup)_i + \beta_3(class)_i + \beta_4(pafr)_i + \beta_5(phisp)_i + \beta_6(pecd)_i + \varepsilon_i$$

Where:

passall: percentage of students passing state-mandated standardized tests, dependent variable

satscor: average SAT score, dependent variable

salteach: average teacher salary

revpup: total revenue per pupil

class: class size, a proxy for student-teacher ratio

pafr, phisp, pecd: percentage of students who are African American, Hispanic, and economically disadvantaged

---

[17] Kiesling, The Relationship of School.

[18] Stiefel, Schwartz, and Ellen, “Disentangling the Racial Test”.

The baseline model investigated the influences of average teacher's salary (H1; proxy for teacher's incentive), total revenue per capita (H2), and class size (H3) on school performance, thus it examined Hypotheses 1-3.

We decided to include SAT score as another separate model even though MacKenzie doubted the reliability of the score. First, SAT scores are still a national standardized test with the potential to reflect students' performance on some level. We regressed the same independent variables on SAT score to see if this would help us interpret the effect of different independent variables. The results of the model using SAT score as a dependent variable were consistent with the model using passing rates on standardized tests as a dependent variable. The addition of the SAT models also made it more clear which independent variables were significant, resolving some discrepancies between the different models that used passing standardized tests as dependent variable.

Furthermore, we augmented our initial model with four other predictors in order to examine Hypotheses 4-7. First, average attendance was used (H4). Second, total number of students within a district was taken as a proxy for district size (H5). Third, average teacher's experience was added (H6). Fourth, percentage of local funding was added (H7).

$$[2a]\ (passall)_i = \beta_0 + \beta_1(salteach)_i + \beta_2(revpup)_i + \beta_3(class)_i + \beta_4(attend)_i + \beta_5(totstud)_i + \beta_6(teachexp)_i + \beta_7(revplocr)_i + \beta_8(pafr)_i + \beta_9(phisp)_i + \beta_{10}(pecd)_i + \varepsilon_i$$

$$[2b]\ (satscor)_i = \beta_0 + \beta_1(salteach)_i + \beta_2(revpup)_i + \beta_3(class)_i + \beta_4(attend)_i + \beta_5(totstud)_i + \beta_6(teachexp)_i + \beta_7(revplocr)_i + \beta_8(pafr)_i + \beta_9(phisp)_i + \beta_{10}(pecd)_i + \varepsilon_i$$

Where:

teachexp: average teacher experience in years

revplocr: percent of revenue from local sources

Results

2005 Data Analysis

We conducted a series of multiple-regression analyses on a sample of 1,039-40 districts for TAKS-based models and 703 districts for SAT-based models. All models passed the model specification test. All the F-statistics reject the null hypothesis that all coefficients are equal to zero (Table 5.1). Therefore, the model specification is sound.

According to adjusted R-squared, our models explain more than 30% of the variation in the percentage of students passing state mandated academic skills tests, with Model 2a having the highest adjusted R-square at 0.410; that is, it explains 41% of the variation in students passing TAKS (Table 5.1).

In order to test for the normality assumption of residuals, we looked at our data in a histogram and through a P-P Plot. Through visual eye-balling both of these tests, we concluded that the normality assumption is met (Figure 1.1).

[Refer to Figure 1.1 "2005 Histogram and P-P Plot of Residuals"]

As some variables displayed high correlation as shown in Table 2.1, a test for multicollinearity using variance inflation factor (VIF) was conducted. As one can see from Table 3.1, the mean VIF for each model is lower than 10, which is the conventional level indicating multicollinearity.[19] Therefore, it is unlikely that the results will have a multicollinearity problem despite hispanic minority and economically disadvantaged variables having high correlation.

[Refer to Table 3.1 "Variance Inflation Factors for Each Model"]

[19] Penn State Eberly College of Science, Detecting Multicollinearity Using Variance Inflation Factors | STAT 501.

Next, we used the Breusch-Pagan test to determine if the homoskedasticity assumption is met in our models. According to Table 4.1, there is heteroskedasticity in each of our models (p-value < 0.05). Therefore, we relaxed the assumption and used robust standard error for our estimations, as suggested by Richards.[20]

[Refer to Table 4.1 "2005 Breusch-Pagan Test for Each Model"]

The direction, magnitude, and statistical significance of the coefficient of each independent variable are demonstrated in Table 5.1.

[Refer to Table 5.1 "2005 Multiple Regression Analysis with Robust Standard Errors"]

2014 Data Analysis

We performed the same analysis with the 2014 data based on a sample size of 777 school districts. All F-statistics reject the null hypothesis that all coefficients are equal to zero (Table 5.2). Therefore, the model specification is sound.

According to the adjusted R-squared values, our models explain more than 45% of variation in the percentage of students passing state mandated standardized tests, or SAT scores. Model 2a has the highest adjusted R-square at 0.682; that is, it explains 68.2% of the variation in average SAT score (Table 5.2).

In order to test for the normality assumption of residuals, we looked at our data in a histogram and through a P-P Plot. Through visual eye-balling both of these tests, we conclude that the normality assumption is met albeit not as evidently as 2005 data (Figure 1.2).

[Refer to Figure 1.2 2014 "Histogram and P-P Plot of Residuals"]

[20] Williams, Heteroskedasticity, [Page 1-16].

Same as the 2005 dataset, a test for multicollinearity using variance inflation factor (VIF) was conducted. As one can see from Table 3.2, the mean VIF for each model is lower than 10 which is the conventional level indicating multicollinearity.[21] Therefore, it was unlikely that the results would have the multicollinearity problem despite hispanic minority and economically disadvantaged variables having high correlation.

[Refer to Table 3.2 “2014 Variance Inflation Factors for Each Model”]

Next, we used the Breusch-Pagan test to determine if the homoskedasticity assumption is met in the models. According to Table 4.2, there is heteroskedasticity in each of our models (p-value < 0.05). Therefore, we relaxed that assumption and used robust standard error for our estimations as suggested by Richards.[22]

[Refer to Table 4.2 “2014 Breusch-Pagan Test for Each Model”]

The direction, magnitude and statistical significance of the coefficient of each independent variable are demonstrated in Table 5.2.

[Refer to Table 5.2 “2014 Multiple Regression Analysis with Robust Standard Errors”]

Differences Between 2005 and 2014 Analysis

We performed a simple difference-of-coefficients test for two independent regressions in order to determine if the magnitude (including direction) of any coefficients had changed significantly from 2005 to 2014.[23] Significant changes in the magnitude of coefficients include:

---

[21] Penn State Eberly College of Science, Detecting Multicollinearity Using Variance Inflation Factors | STAT 501.
[22] Williams, Heteroskedasticity, [Page 1-16].
[23] Clifford C. Clogg, Eva Petkova, & Adamantious Haritou, “Statistical Methods for Comparing Regression Coefficients Between Models,” American Journal of Sociology 100 (1995): 1276.

- Teacher salary had less of a (positive) impact on students passing standardized tests;
- And percent economically disadvantaged students had more of a (negative) impact on both students passing standardized tests and SAT scores.

Interpretation

Revisiting Hypotheses

Based on the results using data from both 2005 and 2014, we can verify whether our hypotheses should be accepted or rejected. The summary of hypotheses and their results are shown in Table 6.

Average teacher's salary has positive and statistically significant coefficients in six out of eight models; therefore, we can conclude that this hypothesis is accepted, and teachers incentive has a positive influence on school district performance. We can summarize the meaning of the coefficients thus: as average teacher salary is increased by $100, students passing standardized tests increases by 10-25%. This might suggest that increasing teacher salary can have a marked impact on how well students do on standardized tests. It may, however, reflect that teachers who work in more affluent school districts get paid more. It is possible that teacher salary, funding, and percent economically disadvantaged students interact in some way.

In 2014, the positive effect of average teacher salary on school performance was significantly diminished. We wonder if this change has anything to do with the increased variation in teacher salary in 2014 (see descriptive statistics) and the high maximum teacher salary.

Total revenue per pupil has negative and statistically significant coefficients in all of our estimations; therefore, we can reject this hypothesis. As school funding per pupil increases by $100, students passing standardized tests decreases by about 9%. The same increase in funding is also associated with a 10-18 point decrease in average SAT score. The direction of causality is noteworthy in this case, though; that is, our finding might not mean that injecting more funding into the school districts deteriorates their performance. It might simply reflect the fact that school districts with worse performances tend to receive more funding in state and federal aid programs.

The coefficient for *average class size* is negative in six out of eight models, most of which are from 2005 data and three of which are statistically significant. The statistical significance lessens when we add other variables to the model. Thus, it might be that the effect of student-teacher ratio overlaps with other variables. Due to ambiguous results, we can neither accept nor reject this hypothesis.

Average daily attendance has a positive and statistically significant coefficient in all four models in which it was included. We thus accept the hypothesis that attendance has a positive influence on school district performance. As average daily attendance increases by 1%, students passing standardized tests increases by about 0.3-0.4%. Across models, attendance also had the greatest single impact on dependent variables. This makes theoretical sense as the connection between attendance rates and school performance is quite direct. Students who are present in class to learn should theoretically perform better on tests.

School district size has three positive coefficients out of four models, only one of which is statistically significant. It can be suggested that school district size might have a positive influence on school district performance, but this is not sufficient to reject the null hypothesis. To be more accurate, more data and analysis are needed.

Average teacher experience has positive coefficients in all models, but is statistically significant in only one model. This suggests that its influence on performance is ambiguous, and we can neither accept nor reject the hypothesis.

Percentage of local funding has positive coefficients in all, but Model 2a of 2014, which is negative and statistically significant. The ambiguous results suggest that we can neither accept nor reject the hypothesis.

markedly from 2005 to 2014. At first, we thought this might be due to the change from TAKS to STAAR. STAAR was said to be more difficult and to put an extra burden on economically disadvantaged.[24]

The United States did undergo a serious economic downturn between 2005 and 2014. This may have increased the number of students who were economically disadvantaged in 2014, which is reflected in the almost 10-point increase in the mean for percent economically disadvantaged student from 2005 to 2014. Students who were economically disadvantaged in 2014 may have also been more severely economically disadvantaged because of the recession. Those experiencing more severe economic conditions could plausibly find it even more difficult to learn at school, thus affecting school performance even more.

Conclusion

Our analysis suggests that some of the most important predictors of school performance are teacher salary, revenue per pupil, and average daily attendance. As teacher salary and attendance increase, so does school performance. As revenue per pupil increases, school performance decreases. While teacher salary and revenue per pupil have complicated causal pathways towards school performance, attendance does not. It seems strikingly clear that one of the best ways to improve school performance--and thus student achievement--is to make sure that kids can actually come to school [reference here...I'll look up later, right now I'm hungry and on a roll]. When we combine this with the fact that the effect that percent economically disadvantaged students has on school performance has increased, we begin to see a clear picture of need in Texas schools. Based on our analysis, we suggest that policies supporting economically disadvantaged students and making sure that they can come to school every day will improve school performance.

[24] Rothstein, The Racial Achievement Gap.

Further analysis is needed to build off of this research model. Although we tried our best to capture what effects test scores, we simply did not have the resources or data to really dive into the issue. Many of the authors we cited from felt that more research needed to be done, since this is a complex issue. We also were left feeling as though more studies, like the Texas Education Agency conducted to get our data, need to be done nationwide. Our analysis would be a lot stronger if we had data from every state and not just Texas. As noted earlier, both texas standardized tests and the SATs changed between 2005 and 2014. While this allowed us to better understand some of the factors that influence school performance, a more direct study could be done on how different types of students and different school inputs show up on differently types of tests. This would be valuable for developing tests free from bias and ones that fairly and effectively measure what they are supposed to.