SMU Data Science Review

Volume 2 | Number 1 Article 1

2019

Tidying and Analysis of the 2014 Texas English II End-of-Course Exam

David Churchman
Southern Methodist University, dchurchman@smu.edu

Abigail Morton Garland
KIPP University Prep, agarland@kippsa.org

Tidying and Analysis of the 2014 Texas English II End-of-Course Exam

David Churchman[1] Abigail Morton Garland[2]

[1] Master of Science in Data Science, Southern Methodist University, Dallas, TX 75275 USA dchurchman@smu.edu
[2] Principal, KIPP University Prep, 239 Stark St, San Antonio, TX 78204 USA agarland@kippsa.org

Abstract. The state of Texas requires all public high school students to take End of Course (EOC) exams. The results of these exams are made nominally public, but in a shape and format that precludes ready analysis. To the extent possible, principles of tidy data will be applied to clean and analyze the publicly released data file for the 2014 English II EOC exam, providing insights into the EOC program and a case for better public data from the Texas Education Administration (TEA).

1 Introduction

Since 1980, high school students in Texas have been required to pass high-stakes standardized tests in order to graduate. Successive iterations of have been developed with the intention of increasing rigor for students and accountability for students. The federal No Child Left Behind Act of 2001 increased the stakes for the exams, tying large amounts of funding to student performance. The number of exams high school students were required to take and pass grew in number and rigor until 2012, when it was announced that high school students would need to pass fifteen EOC assessments. Public backlash from parents, teachers and students pressured the legislature to trim the exam requirements, and in 2013, they passed House Bill 5 (HB5), which mandated five end-of-course assessments for graduation. [23]

Besides reducing the amount of standardized testing to relieve public pressure, the stated intention of HB5 was to ensure that high school students in Texas were "college ready." HB5 defines students as college ready if they perform at certain benchmarks on tests such as the PSAT, SAT or ACT. The Texas Education Agency publishes data every year on performance on the EOC, ACT and SAT exam at Texas public schools, however it does not provide any descriptive analysis showing that EOCs necessarily prepare students for the ACT or SAT. Schur's 2015 thesis [22] found some limited qualitative and quantitative evidence that the SAT scores of the first graduating class under HB5 dropped, though suggested further analysis was warranted once more graduating classes were available. Other analyses have found broader impacts of HB5, such as a dramatic reduction in the number of students taking Algebra II [25] or the

perpetuation of educational inequality among students of color [26]. One study found that the number of students at a high school had a significant impact on EOC results [11].

Students that started high school at the time HB went into effect only recently started to graduate. The graduating class of 2016 were in tenth grade the first year under HB5, and data related to their EOC and SAT/ACT performance have been released recently by the TEA. The format of this data–especially the EOC data–precludes ready analysis. Hadley and Wickham's principles of "tidy data"[28] is applied to clean the messy EOC data. Cleaning is done with reproducibility in mind, so as more EOC data is released, cleaning can be expedited. Visualization and analysis is done on the cleaned EOC data as an initial evaluation of whether the EOC program meets the stated intention of preparing students that are college ready.

2 Overview of the EOC Exams

2.1 Exams

Public school students in the state of Texas take the State of Texas Assessments of Academic Readiness (STAAR) starting in the third grade. At the high school level, the tests are called End of Course Exams (EOC) or STAAR-EOC exams; the terms EOC and STAAR are generally used interchangeably at the high school level. There are five EOC exams Texas high school students are required to pass to be eligible to graduate: Algebra I, Biology, English I, English II and U.S. History. Because the exams are required for graduation, they are considered "high-stakes." [15] There is no requirement for what grade students must be in to take the exams, and in fact, there is a trend where Algebra I is increasingly taken by students in the eighth-grade rather than in high school. While there is no requirement, the majority of students take Biology and English I in their ninth-grade year, and Algebra I if they did not already pass it the previous year, making ninth-grade the most important for determining whether students will be able to graduate based on their EOC results. Subsequently, most students take English II in tenth-grade and U.S. History in eleventh.

Students are required to take the exams at the "end of course," meaning in April or May of the year they take the eponymous course of the exam. If a student fails to pass an EOC on the first attempt, they are then allowed to retake the EOC as many times as necessary to pass. There are EOCs proctored in June and December of each year in addition to the spring administration for these re-testers. Released EOC scores will always include these re-testers, who are more likely to score lower than those students taking the exams for the first time; students who struggled with the material once are more likely to struggle each subsequent attempt. There is no disaggregation between first time testers and re-testers in reporting, nor is there information from which this might be inferred, such as the grade levels of the testers–twelfth-graders who are taking the English I exam are more likely to have taken it several times before than ninth-graders and are likely to perform very differently.

The Texas Education Agency (TEA) has published aggregated STAAR data from the 2011-2012 to 2017-2018 school years. There options for data files aggregated at the region, district and campus level. Hadley-Wickham define tidy data as having each row correspond to an individual observation. In the case of testing data, a single observation would be an individual student's performance. The Family Education Rights and Privacy Act (FERPA) protects students from educational data being published that could be linked to their individualized performance, which is why the TEA does not release individual student data. The most granular level of data publicly available is the campus-level. The TEA could maintain compliance with FERPA by publishing anonymized, student-level data, removing identifying information and masking data that might be used to infer a student's identity. However, the TEA did not choose this approach, so the campus-level data, as the most granular, will be used as the observational unit.

An aggregation of students into a campus will not provide robust information on every single student within each campus. One unfortunate reality does make each campus closely representative of the students at that campus: Despite efforts in the Civil Rights era, public schools are still extremely segregated by racial,socioeconomic and linguistic factors. [21] There are over 2,000 campuses contained within TEA's released EOC data, and due to extreme segregation, effects of race, socioeconomic status and language, among others, can still be explored.

2.2 Scoring

The method for scoring of the STAAR tests has shifted in that timeframe, frustrating longitudinal analysis. Across all the years, EOC tests were given a raw score, which was simply the number of correct answers for a student [12]. Most EOC questions are multiple choice with a single correct answer choice, making a raw score relatively straightforward to calculate. The Algebra I test has between four and six questions that have an open-ended numeric response which can be entered into a specialized grid and scored by machine. The English I and II EOCs are the only two with a human-scored question–a short written composition. A raw EOC score is converted to a scale score using a conversion table, which changes each year. Typically, the scale scores are from 0 to approximately 6100, but the max scale score can vary from even the December to May administrations. Different scale scores are assigned to descriptive performance bands. From the 2011-2012 school year to the 2015-2016 school years, there were three performance bands students could fall into for each test: Level I Unsatisfactory, Level II Satisfactory, and Level III Advanced. Levels II and III were considered passing, where Level I performance required re-testing for a student to be eligible to graduate.

The scale scores necessary for Level II and III steadily rose in this time frame, but cohorts of students were grandfathered into older standards. Students were entered into cohorts based on the year they first took any EOC exam, which was typically determined by when students took Algebra I: eighth or ninth grade.

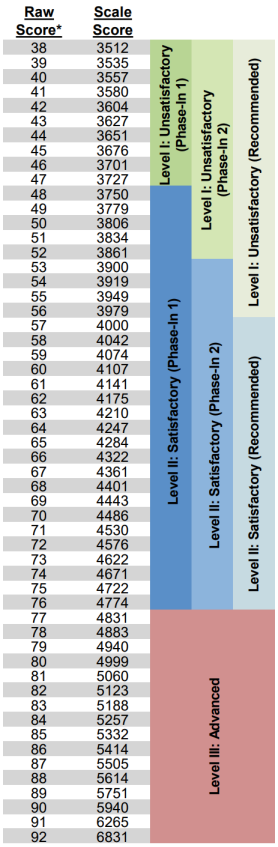

Fig. 1. In the Spring 2014 English II administration, different students taking the same test needed either 48 or 53 answers correct to achieve Level II, depending on their cohort. It was also "recommended" that students attempt to get 57 questions correct, though this was not an official criteria to be called Level II so it is unclear why any school would heed this recommendation.[12]

This had the effect that two students in the same grade level taking the same exam and giving the exact same answers could fall into different performance bands. A student in the Level I performance band would have to retake the test in order to graduate while the Level II student would not. Figure 1 is an example of the type of conversion table published for every test administration, amounting to three per EOC per year. In the 2013-2014 school year, a fourth performance band was added and all of the bands were renamed. They were no longer "Level I," "Level II" and "Level III." Now, students receive a designation of "Did not meet," "Approaches," "Meets," or "Masters," referring to whether students met grade level expectations on the exams or not. A designation of "Approaches," "Meets" or "Masters" all are now considered passing, and only students who achieve "Did not meet" are required to retake the exam.

If EOC data from before 2016 and after 2016 were both to be compared, there would need to be a decision about how to compare the performance bands across the years. The most obvious comparison is between "Level I: Unsatisfactory" and "Did not meet," both of which were the failing conditions for their respective exams, and the only score with high stakes for students. Students do not submit EOC scores on college or job applications; the largest direct effect on a student is whether the exam becomes a barrier to graduation. It is less clear how to compare the other performance bands. The name, *Approaches* acknowledges that passing these exams at the lowest level is not necessary to demonstrate college readiness, so it would be helpful to have a meaningful comparison of students that are demonstrating a higher level of performance. Is *Level II: Satisfactory* analogous to *Approaches* or to *Meets* or to both *Approaches* and *Meets*, leaving *Level III: Advanced* to correspond to *Masters*?

Table 1. Spring EOC Passing Rates since HB5

Year	**Eng I**	**Eng II**	**Alg I**	**Biol**	**US His**
2014	62%	66%	81%	91%	92%
2015	63%	66%	81%	92%	91%
2016	63%	66%	81%	91%	94%
2017	60%	62%	82%	85%	92%
2018	60%	66%	83%	87%	92%
Average	*62%*	*65%*	*82%*	*89%*	*92%*

Besides being inconsistent in scoring from year to year, or even within the same year as a December score table can be different than the spring, the tests have widely different passing rates from each other. Table 1 shows the spring passing rates for EOCs since HB5 passed and the tests were changed to adhere to HB5.[4] The rates have been very consistent since 2014, where Algebra I is always between 81-83% passing, Biology and U.S. History are between 91-94% passing, and the English EOCs are in between 60-66% passing. The chief stake of the tests are whether students are able to pass and graduate, so effectively,

the most difficult test to pass is the most important to look at for any analysis. This makes the English tests of much greater interest than the other three, as they are most likely to be the assessment that keeps a student from graduation.

3 College Ready

The stated intention of HB5 was to create more college ready students; however, the law did not define college readiness. In 2014, before HB5 was passed, the Texas Education Agency (TEA) published a rosy report on the college readiness of graduating Texans [10], in which they showed a steady improvement of college readiness in both English Language Arts and Mathematics, as seen in figure 2. In this report, a student could be recognized as college ready via either their TAKS score-–TAKS was the predecessor to EOC–SAT score or ACT score. The *or* condition meant that students could be defined college ready based solely on their TAKS score, which is a self-referential loop. Were students getting more prepared for college or was Texas making it easier to perform on their state tests? There is no suggestion in TEA's report that an indicator outside of the test itself should be used to measure the success of the test at helping to create college ready students. Similar to the EOCs, the TEA publishes campus-level

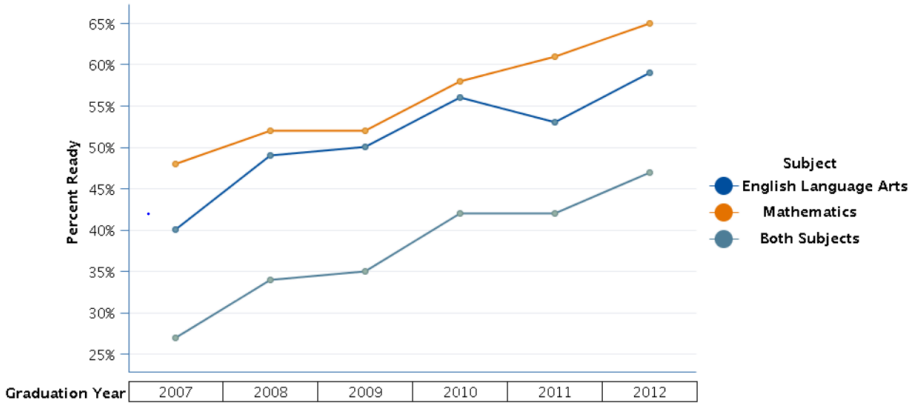

Fig. 2. Excerpted from a TEA report [10] showing a dramatic rise in "College Readiness" for students in both English Language Arts and Math, as defined by performance on the TAKS, SAT or ACT.

SAT and ACT data each year. The TEA has defined "critical rates" for the SAT and ACT that qualify students as "college ready" and report what percent of each campus has achieved that rate on at least one assessment. However, the SAT/ACT data is released by TEA at a lag from the EOC data. The most recent campus level data, from the class of 2016, was released in January of 2019. The

class of 2016 took the English II EOC the first year of the HB5 compliant EOCs in 2014. Previously, the U.S. History tests could have been compared for the class of 2015, but the passing rates on the U.S. History EOC make for a less informative analysis.

In addition to the SAT or ACT, students are designated "college ready" based on advanced courses. Many Texas high schools enroll students in Advance Placement (AP) and International Baccalaureate (IB) courses, whose intention is to give students an opportunity to earn college credit while in high school and are considered advanced. An AP score of 3 out of a possible 5 is considered equivalent to passing a college course. Similarly, an IB score of 4 out of 7 is considered a college equivalent. The TEA published aggregated AP and IB results for the 2011-2012 through 2015-2016 school years. Whereas almost all high schools have at least some students participate in the SAT or ACT tests, many fewer schools have opportunities for AP or IB.

One aspect of each of these measures of college readiness is that they are all voluntary on the part of the student. A school could conceivably produce college ready students that do not take the ACT, SAT or advanced classes. There are fees associated with all of these programs, though waivers can be obtained for students designated economically disadvantaged and counselors knowledgeable enough to help students get those waivers. One classic statistical study attempting to find the cause of disparities in mean SAT scores between different states [24], found that the differences were not attributable to the inherently greater education in certain states but almost entirely by the participation rate of students in each state. The TEA definition of college ready aside, a more minimal requirement for entry to and subsequent success in college is whether a student graduates from high school. Graduating from high school is a necessary–though most would argue–not sufficient, condition for attending college.

4 Choice of the 2014 English II EOC

The recent release of the class of 2016 SAT/ACT results by campus makes possible the comparison to the 2014 English II EOC, which most students in the class of 2016 took when they were in tenth grade. These students were the first students to take the English II EOC under HB5, making them an interesting class to explore. Though the campus-level nature of the data for both the EOC and the SAT/ACT precludes a true cohort analysis, comparisons can still be made, and if effects are strong enough, they may still be seen. The analysis will focus on this graduating class for these two exams, and will pull in data from a third data file for the graduation rate for this same class. The three data reports to be analyzed are:

- 2014 English II EOC
- 2016 SAT/ACT Critical Rate
- 2016 Graduation results

5 Student Sub-populations

Based on requirements from No Child Left Behind (NCLB), TEA's data is disaggregated across many different sub-populations of students in order to test whether any group is being educated at a different level. This creates a rich set of potential features to analyze, but also such a large number of features that an understanding of the sub-populations reported on is necessary. For example, the first sub-population listed on every report is sex. Texas recognizes "male" and "female" as options for sex in its data reporting, though also records the number of students for which no sex information is provided. Most of the students for whom no sex information is provided are probably the result of a reporting error, but it is also likely that there are an increasing number of students and instructors uncomfortable with the binary sex designations, [17] opting instead to omit sex information for a student, making the "no sex information students" less clear in the analysis. Every subsequent sub-population described has the "no information," and depending on the sub-population, could have different meanings.

Analyzing the ethnicity sub-populations presents similar difficulties as gender. The seven recognized ethnicities, in addition to "no information," are Hispanic/Latino, American Indian/Alaska Native, Asian, Black or African American, Native Hawaiian or Pacific Islanders, White, and Two or More Races.[5] Race or ethnicity is self-reported by students, and there are anecdotal reports that members of an ethnicity that experiences regular discrimination will not accurately self-report [27], fearing it will cause further discrimination. This is also a potential issue for the self-reporting of migrant status.

Student data is also disaggregated by whether a student is economically disadvantaged. Socio-economic status has been shown to be a significant predictor on Texas assessments in the past. [20] Families apply each year for this program by reporting their income. If their family income is 1.3 [2] times the federal poverty line for their family size or lower, they qualify for free lunch, or 1.85 times or lower for reduced price lunches. If a student qualifies for either free or reduced lunch based on their family's income, they are considered economically disadvantaged. If a family does not apply for free or reduced lunch, they can still be labeled as economically disadvantaged if they received any kind of federal assistance. [7] EOC results are reported for the grouped "economically disadvantaged" sub-population, as well as broken out for "free lunch," "reduced lunch" and students that qualified in another way.

If a school's enrollment is forty percent or more economically disadvantaged, it is eligible for Title I grants [9] and all students are labeled as being at a Title I school, whether or not they are economically disadvantaged themselves. Schools that do not meet the forty percent threshold can still qualify for Title I funds for students that meet specific requirements, such as being homeless or neglected. At the campus level, Title I is essentially a binary variable of either all students or no students based on whether the school crossed the 40% threshold.

Data is also disaggregated for students who are Limited English Proficient (LEP). Students are designated as LEP based on their performance on the Texas

English Language Proficiency Assessment System (TELPAS) [6], which is administered to any K-12 student that teachers or administrators believe may be non-native speakers of English, and then every year by students previously labeled as LEP. LEP-designated students are eligible for additional services, with the goal that they are eventually exited from LEP status based on their TELPAS score. Then, they are monitored for two years to insure they maintain this level of English proficiency and are no longer LEP. EOC scores are reported for LEP, first year monitored and second-year monitored students. Similarly, students are assessed for special education status [3], though there is no single assessment system as there is for LEP status. Students can qualify for special education based on a wide range of physical, emotional, behavioral or psychological disabilities.

Students are also labeled as "at risk" or "not at risk." There are thirteen ways a student can be labeled as at risk [8], including failing a grade, a STAAR test, or being expelled from a school. These sub-populations with several others, including gifted status and whether a student participates in Career Technical Education (CTE), total to 64 sub-population categories that are reported by TEA. This is already a large feature set for many tools of analysis, but each variable reported for a campus is generally also disaggregated by the 64 sub-populations, so it multiplies into thousands of features in a single report.

6 Data Preparation

6.1 Width vs. Length

TEA publishes a data report as a '.dat' file for each school year for each EOC, making a total of 35 reports for the seven years of STAAR testing from 2011 to 2018. The dimensions of each report vary, the most significant difference being the change from three performance bands to four in the 2017 report, which is then multiplies by each sub-population making for hundreds of extra features. The large number of sub-populations means a very large number of features in each report. The 2014 English II report has 1,943 features for each campus. There are six identifying features for the campus, including the year, campus number, region and district numbers and campus name. There is the number of total tests submitted, students reported absent and tests not submitted for a reason other than absence reported both as a number and a percent for six more features. The remaining 1,931 features are reported for all 64 sub-population categories described above.

The total number of students and average scale scores are reported for each sub-population for 128 more features. The number and percent of students is also reported for each of the four performance bands, making for 512 more features, and the majority of the features comes from reporting both the average number correct and average percent correct for each content category, of which there are six for English, contributing 768 features. Having 1,943 features for a single EOC for a single year is impractical. There are only 2,060 campuses in the 2014 English II data, so there are almost as many features as observations, which would likely lead to poor models in the analysis. Table 2 presents a small example of the

Table 2. Wide and Untidy Example Table

Campus	Black Tested	White Tested	Hisp Tested	Black Passed	White Passed	Hisp Passed
A	2	4	1	1	1	1
B	5	3	7	2	2	2

unwieldy format of the EOC data file. With only three sub-populations and two variables per campus: number tested, number passed, a wide table that is difficult to analyze or make comparisons on is generated. It is difficult to tell which campus performed better and in which sub-populations. Hadley-Wickham define tidy data as each variable being sorted into a single column. Table 2 violates this rule, as the variable "number tested" is stored across three columns, one for each sub-population. Similarly, "number passed." A variable stored across multiple columns is not only difficult to read, but makes aggregations, visualizations and analyses computationally inefficient. Table 3 rearranges the same information in

Table 3. Long and Tidy Example Table

Campus	Group	Number Tested	Number Passed
A	Black	2	1
A	White	4	1
A	Hispanic	1	1
B	Black	5	2
B	White	3	2
B	Hispanic	7	2

table 2 into a tidy, long format. It is now possible to quickly compare the number tested and passed in each group of students at each campus. If a new column of "percent passed" wished to be calculated for each group, it would be a simple matter of dividing the elements of two columns. In the untidy format, five new columns would be needed, requiring five times as much code to complete. This general principle of taking the very wide EOC data and reshaping it to a tidier and longer format was the largest portion of data preparation.

Data preparation was performed using the R programming language, the code for which can be found at the author's GitHub repository. [13]. One advantage of R is creates a reproducible set of steps for other researchers to build upon, and for others to be able to check the assumptions and calculations used in this paper. [14] As much as possible, code was written to be generalized to other EOCs, though the shifting nature of the data files and scoring categories mean adjustments would be necessary.

6.2 Corrupted Data Files

The 2014 EOC data files were released with an error in the formatting of its header row, meaning that standard reading techniques such as `read.delim()` or `read.csv()` could not accurately determine the dimensions of the data but instead read the header onto two separate rows of unequal length. A function was written to correct the header and adjust the dimensions accordingly.

As discussed above, the EOC data also is published in a very wide format, where each row is a single campus, and there are thousands of variables for each campus, many redundant. The SAT/ACT data, however, is published in a tidy, long format, where there are multiple rows for each campus, where each row represents a different student sub-population. As these sub-population reporting categories are mandated by NCLB, they are the same between EOC data and SAT/ACT data. The largest difference is that the EOC data reports exact numbers for all sub-populations and reports for how many students it is missing information for each sub-population, where the SAT data masks numbers if they are small enough that individual student results might be inferred and does not report on missing information.

The tidy, long format used by the SAT/ACT data is preferable, so functions were written to reshape the English II EOC to match the SAT/ACT format. This meant finding the corresponding columns across the thousands of columns in the EOC data and transforming them into rows that could be joined with the SAT data. The graduation data is published by the TEA in a similar wide format to that of the EOCs, which was also reshaped to be joined to the SAT/ACT data.

Both the graduation and SAT/EOC data use masking, frequently rounding values to the next 25 and reporting the number in the format "<125". These were converted to numeric, stripping the "<" symbols, but it is important to note going forward that these do not represent exact numbers.

7 Visualization and Analysis

7.1 Student makeup

In order to make effective comparisons among the three data sets, one must assume that the students comprising the class of 2016 for each campus are essentially the same students as those that took the English II EOC in 2014. The campus level nature of the data makes it impossible to check this definitively, but with evidence to support this assumption, some comparison can be justified. Figure 3 plots the reported number of students for each campus in pair plots comparing the English II table, the reported graduating class size, and the number of students in the graduating class reported by the SAT/ACT. One would expect the class size on the SAT/ACT and the graduation report to be the same, as they were both reported the same year by the TEA. The plots show that they are tightly correlated, though not identical.

The variation between EOC totals and the other two reports is more explainable, as the numbers of students at a campus could genuinely change over

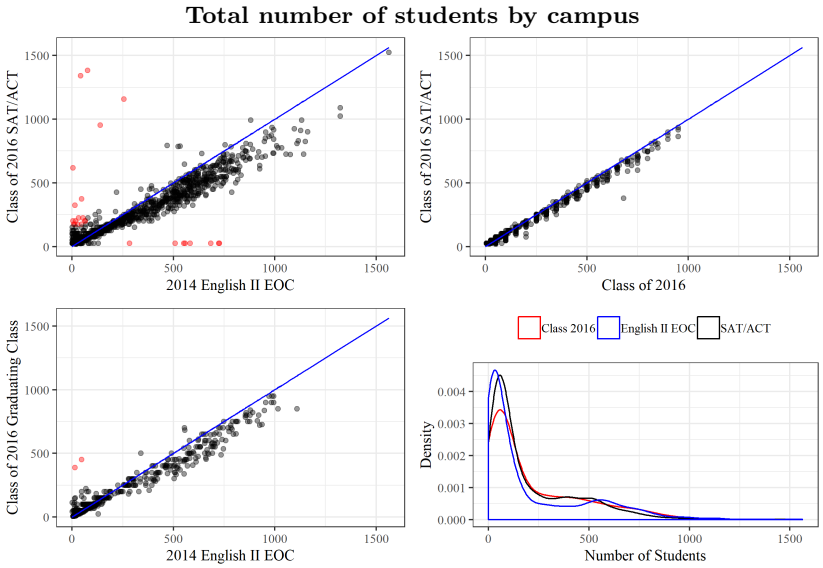

Fig. 3. There is strong correlation between the number of graduates reported by SAT/ACT in 2016 compared to how many took the English II EOC at each campus. A blue reference line is plotted for $y = x$, and the number of points below the line in the left two plots is suggestive of dropouts.

the course of two years. Students could move between schools, making class size larger or smaller. The blue line is plotted at $y = x$ on each plot rather than a regression line to help illustrate the differences between each variable. If students were simply shifting between schools between 2014 and 2016, one would expect a random scatter of points around the blue line, with roughly the same number of schools growing in number of students as shrinking. However, what we see is that there are fewer students in the class of 2016 than when they took the English II EOC in 2014, which suggests that students are not just moving between schools, but dropping out from Texas schools altogether. In the same period, the population of Texas grew by close to one million people [1], making it likely that students are not leaving Texas high schools for high schools outside of Texas, but are leaving high school all together.

In figure 3, the most dramatic outliers are colored red. These outliers were identified as schools whose populations were either tripled or divided by a factor of three in the two year period, with an initial or final number of students above the threshold of 150 students. These numbers were chosen visually to capture the schools most radically different from the other. Of the almost 2,000 schools which have data for all three sets, it only applies to 34 schools. These schools have undergone radical transformations and can not be thought to represent the same populations of students, so these outliers are removed from subsequent analysis, leaving 1,928 schools that have data in all three reports.

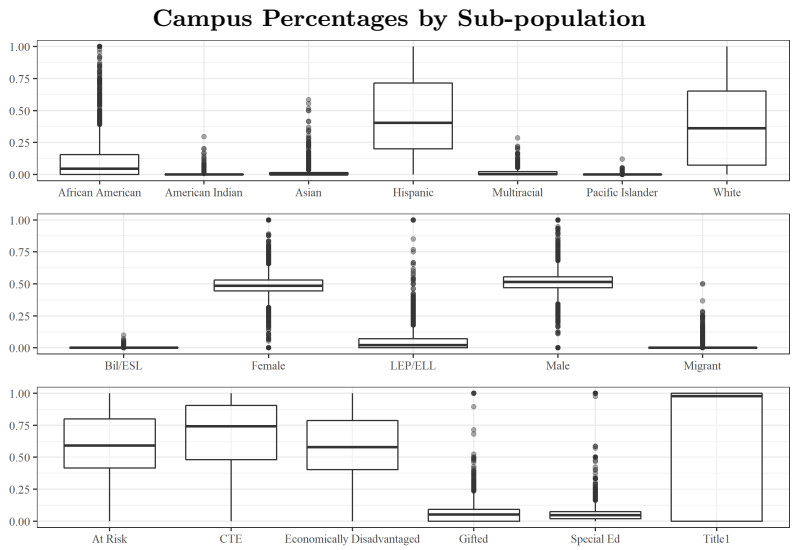

Fig. 4. Boxplots for each sub-populations representation at campuses, as reported in the 2014 English II EOC data.

The masking of data in the graduation and SAT/ACT report did not affect total numbers of students dramatically, as total population was rounded but generally included. The masking on sub-populations was more dramatic, and comparisons in the quantities of populations were unproductive. Figure **??** gives a box plot of the percentage of each campus by sub-population according to the English II data report, which was the only one that did not mask the raw number of students within each population.

According to the EOC data, students in Texas predominately identify as hispanic and white, with a significant minority identifying as African American. There are very few campuses with significant populations of the other races or ethnicities. Due to these very low numbers, the Asian, American Indian, Multiracial and Pacific Islander results were combined into a single "Other Race and Ethnicity" category.

Other than ethnicity and race, most schools are half female and half male, though there are all-boys and all-girls schools and schools that are skewed more heavily female or male. The Title I designation is always either 100% or 0%, as discussed above. Just more than half of Texas public schools qualify as Title 1. Most schools have a small percentage of students who are designated as gifted or as needing special education. The "Bil/ESL" designation appears to have been almost completely replaced by "LEP/ELL", so it was dropped from further analysis. There are very few schools with any migrant population, with some outliers.

7.2 Correlations between Assessments

Figure 5 plots Spearman Correlation Coefficients for each pairwise comparison within the cleaned and combined data set, providing a rich set of insights. Spearman was used rather than Pearson' correlation coefficient, which measures the linearity of relationships, as it measures the strength of any monotonic relationship. The grid is organized with the fourteen remaining sub-populations as the first columns on the left, and the five measures of student success as the last columns. These last five columns are:

- The percent of students which passed the English II EOC at each campus in 2014
- The percent of students that achieved "advanced on the English II EOC
- The percent of students that graduated in 2016 from each campus
- The percent of students that participated in the SAT or ACT in the class of 2016
- The percent of students that achieved the "college ready" critical scores on the SAT or ACT in the class of 2016

Correlation Plot of Sub-populations and Results

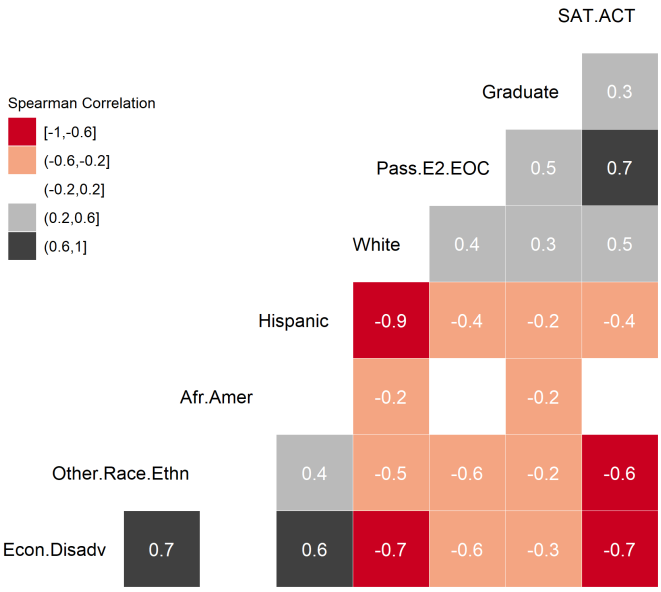

Fig. 5. Spearman correlation coefficients are plotted for each pairwise comparison of sub-population and result. Red indicates a negative monotonic correlation and black, a positive monotonic correlation. The blue line divides student sub-populations and student results.

Starting in the top right corner of figure 5, there does appear to be some correlation between the different measures of student success. The strongest correlation is between the percentage of students at each campus that passed the English II EOC in 2014 and the percent that met the critical rate on the SAT or ACT in the class of 2016. It is surprising that this relationship is stronger than the correlation between percentage of students that achieved "advanced" on the English II compared to the SAT/ACT. Figure 6 explores these relationships more closely.

Figure 6 reveals that there does appear to be a mostly linear correlation for both "passed" and "advanced" and the SAT/ACT. It appears to be more difficult to achieve the critical rate on the SAT/ACT than to pass the English II, and then more difficult even still to get advanced on the English II EOC. If the ultimate goal is to produced students that are able to achieve the critical rate on the SAT/ACT and thereby be identified as "college ready," there appears to be a missing "sweet spot" between "passing" and "advanced" on the EOC. In 2016, an extra performance band was added to the EOCs, which may serve this role. The left plot of figure 6 also reveals the bleak results for both the EOC and the SAT/ACT at Texas schools. Each point represents a campus, and the points are centered around 50% of students at each campus passing the EOC, and less than 25% achieving the critical rate on the SAT or ACT.

Returning to the top right of figure 5, it is surprising there is not a stronger relationship between the percent of students that participated in taking the SAT/ACT as compared to achieved the critical rate. Figure 7 makes clear why this might be. One, there are forty schools reporting higher a percentage of students achieving the critical rate than participating, which is impossible. These outliers are most likely due to reporting errors and could be removed, but a much larger group of outliers, shaded green, still prevents any correlation.

These outliers represent schools that have a very high percentage of students, close to or at 100%, which sat the SAT or ACT, but do not perform at the critical level. The authors of this paper work at such a high school. The school serves predominately low-income, at risk, Hispanic students with the explicit purpose of getting them to college. One barrier is admissions tests, so all students are required to take them, regardless of their ability. These green outlier schools in figure 7 most likely represent similar high schools that require all students to take the assessments, regardless of their ability.

Next, figure 8 investigates the correlation between the percentage of students passing the English II EOC at each campus and the percentage graduating two years later. With the exception of five campuses with 0% of students graduating, it does appear that the percentage of students passing the English II EOC serves as a floor for the percentage of students that will graduate. As a rule, many more students seem to graduate than pass the English II EOC. This means that students that did not pass the EOC on the first attempt must pass it in the next two years, as they are allowed to retake it many times as necessary. There does appear to be a stark cut-off where schools that have fewer than 25% of students passing the EOC have dramatically lower percentages of students graduating.

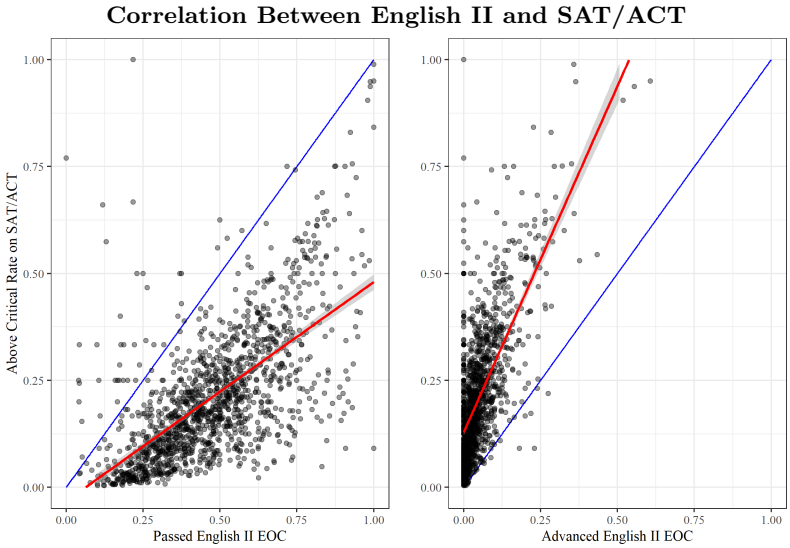

Fig. 6. The blue line represents a one-to-correlation, and the red line represents the line of best fit. There is a correlation between both passing and getting advanced on the English II EOC and achieving the critical rate on the SAT or ACT. There is evidence that achieving the critical rate on the SAT or ACT is somewhere between the difficulty of "passing" and "advanced" on the English II EOC.

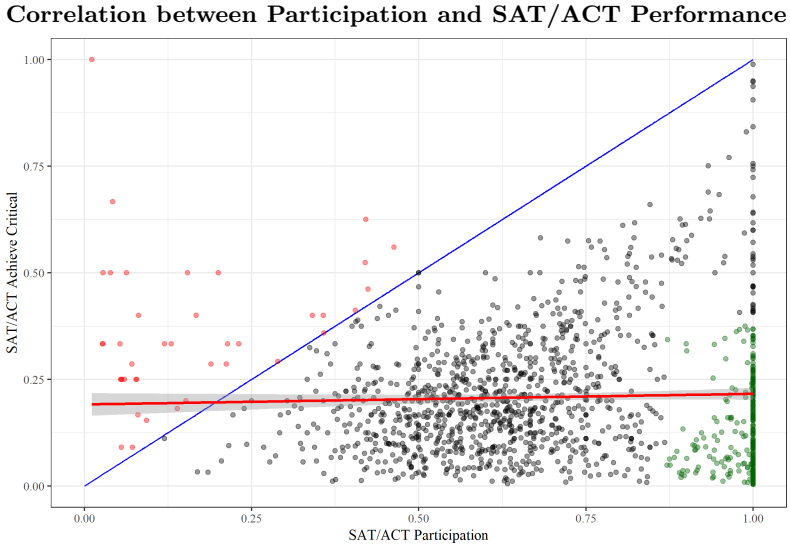

Fig. 7. There are 40 schools that reported a higher achievement rate than participation rate, marked by red points. There are also a large number of schools with high participation on the ACT/SAT, but very low achievement.

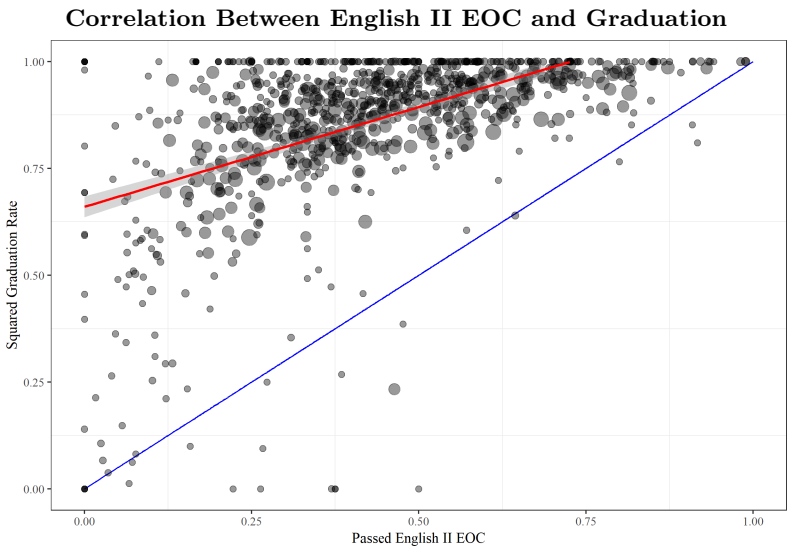

Fig. 8. There is a weak correlation between the percentage of students that pass the English II EOC and the percentage that graduate. Points are sized proportionally to the number of students in the class.

7.3 Correlations in Sub-populations with Results

This section will analyze the area to the left of the blue line in figure 5, looking at student sub-populations. First, sub-populations that did not appear to have an effect on results. The percentage of students that are female or male did not have any relationship to anything else significant; most schools are evenly divided, as seen in figure 4. The percentage of students participating in Career and Technical Education did not register. The only sub-populations that were positively correlated with results are "white" and "gifted." Both were weakly correlated with students performing better on both the English II EOC and the SAT or ACT. The gifted correlation suggests that schools with gifted programs, perform better than those that do not. There are negative correlations with assessment scores and the percentage of students at schools that are any race besides white.

The most strongly negative correlations with results were for schools with large numbers of students labeled "at risk" and for schools with large numbers of students labeled as "economically disadvantaged," based on their qualification for the federal free-lunch program. "Economically disadvantaged" is also strongly negatively correlated with "white." In general, there is evidence to suggest that schools with more white students and fewer students from low-income families tend to perform better on the English II EOC. This correlation is even stronger for the percentage of students that achieved the critical rate for the SAT and ACT.

8 Ethics

This project has attempted to do a rough cohort analysis of students using publicly available data. The data reports do not have any personally identifying information about individual students, and numbers were masked in the graduation and SAT/ACT data. However, student numbers were not masked in the EOC data. Sub-populations with small numbers of students at a certain campus may allow for narrowing down individual student EOC results, and then infer who those same students are for the SAT/ACT data. FERPA protects students from their individualized results being published, and the data should not be abused in this way.

There is also a long history of using various forms of "unbiased" assessments to attempt to prove differences in intelligence between different races and ethnic groups. [16] These results have been widely and repeatedly debunked. The strongest research suggests that the fundamental intellectual capacity of humanity has not changed in 75,000 years. [18] Any differences in achievement on standardized tests between racial or ethnic groups is evidence of a difference in access to high quality education, not potential intellectual capacity of students in different groups.

Some have convincingly argued that the entire premise of standardized testing is harmful to students and schools. [19] Standardized tests claim to be objective measures of achievement, but time and again, it has been shown that they reveal more about the implicit biases of the testers than the actual educational achievement of the tested. Tests like the ACT or SAT not particularly predictive of college readiness nor really any other measure of success. In this study, the results on one test were correlated with another test, which is an Ourobos of self-justification. Despite their well-researched flaws, the results of these tests are taken very seriously, justifying labeling schools and students as *failing*, and fostering an unhealthy, competitive culture.

9 Conclusions and Further Work

For this project, `R` scripts were written that would work to tidy any of the 2012-2016 EOC exams and perform similar analysis to that seen in this project, but the format changed dramatically in the 2017 results, so some adjustment would still be needed for subsequent schoolyears. The results for assessments in Texas are putatively public. However, in practice, the results are posted in corrupted and untidy formats that make analysis inaccessible without significant processing.

After processing the 2014 English II EOC data, results were suggestive that there is some alignment between success on the English II EOC and students achieving "college ready" marks on the SAT or ACT, though many fewer students are able to achieve the latter than the former. There is also suggestive evidence that Texas schools with higher percentages of white and higher-income students have an easier time getting students to pass the EOCs, and even more-so the SAT and ACT. The only student sub-population, besides "gifted" with

a positive correlation with success on the English II EOC, graduation, and the SAT/ACT was white students, and the strongest negative correlation was with economically disadvantaged students, suggesting evidence that Texas is providing unequal education to low-income, non-white students, as illustrated in figure 9.

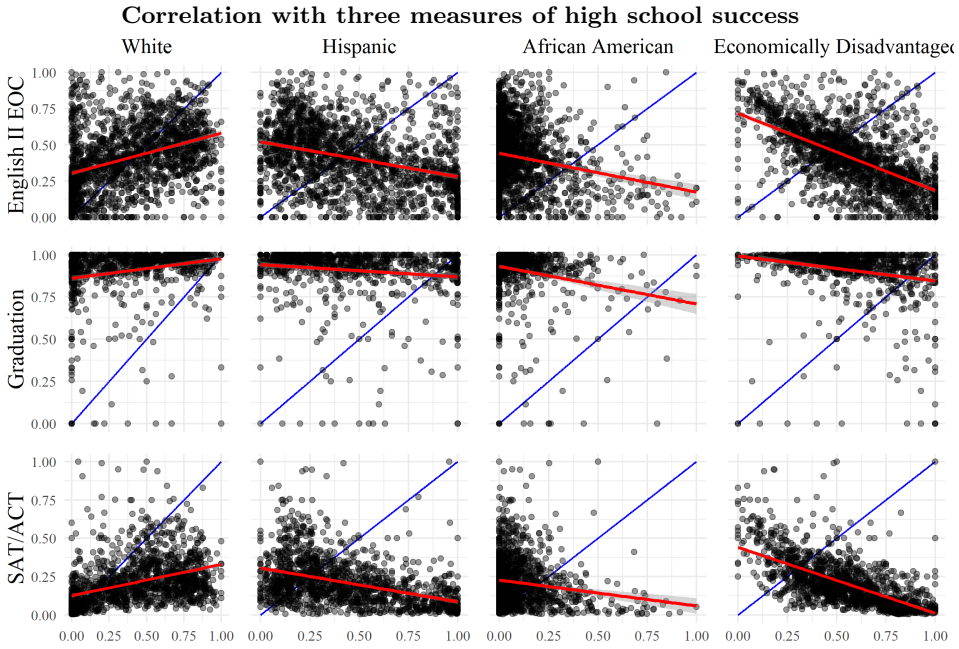

Fig. 9. The first row is each Texas campus' 2014 pass rate on the English II EOC. The second row is the 2016 graduation rate. The third row is the percent of each campus in 2016 that achieved "college ready" on the SAT or ACT. The only sub-population, besides "gifted" with positive correlations with success on these three measures was "white." The strongest negative correlation was with "economically disadvantaged."

The campus-level nature of the data make it difficult to prove the correlations are true, but they provide suggestive results. No Child Left Behind was meant to insure equal access to a strong education for all students, but the results in Texas suggest that more non-white and low-income students are currently left behind in achievement. For more conclusive analysis, the TEA should provide student-level data, with identifying student information masked. The TEA should begin releasing all data reports hewing closely to tidy data conventions.

With tidier data and data that allows for cohort analysis, further work should be done to see if the English II EOC is serving its stated purpose of preparing students to be college ready. Further comparisons should also be made to other

assessments of college readiness, such as Advanced Placement and International Baccalaureate exams. As students that have gone through the full testing regime under HB5 graduate and attend college, actual outcomes in college, including matriculation, first year completion and eventual graduation should be analyzed.

References

1. Census data. https://www.census.gov/.
2. Child nutrition programs: Income eligibility guidelines. https://www.federalregister.gov/documents/2018/05/08/2018-09679/child-nutrition-programs-income-eligibility-guidelines.
3. Special education rules and regulations. http://framework.esc18.net/Documents/Side_by_Side.pdf.
4. Staar statewide summary reports. https://tea.texas.gov/staar/rpt/sum/.
5. Student subgroup. https://www.edglossary.org/student-subgroup/.
6. Telpas resources. https://tea.texas.gov/student.assessment/ell/telpas/.
7. C054 economic-disadvantage-code. http://ritter.tea.state.tx.us/peims/standards/1314/c054.html, Nov 2018.
8. E0919 at-risk-indicator-code. http://ritter.tea.state.tx.us/peims/standards/1314/e0919.html, Oct 2018.
9. Improving basic programs operated by local educational agencies (title i, part a). https://www2.ed.gov/programs/titleiparta/index.html, Oct 2018.
10. Texas public high school college ready graduates by region. http://www.texaseducationinfo.org/TEA.TpeirPortal.Web/Reports/College%20Ready%20Graduates_Region.pdf, Oct 2018.
11. John H Bishop, Ferran Mane, Michael Bishop, Joan Moriarty, Richard Murnane, and Laurence Steinberg. The role of end-of-course exams and minimum competency exams in standards-based reforms. *Brookings papers on education policy*, (4):267–345, 2001.
12. Eli Cadena. Staar raw score conversion tables. https://tea.texas.gov/student.assessment/staar/convtables/, Sep 2018.
13. David Churchman. David churchman's github repository. https://github.com/dpchurchman/E2EOC2014, March 2019.
14. Christopher Gandrud. *Reproducible research with R and R studio*. Chapman and Hall/CRC, 2016.
15. Emily Germain and Michael Barnes. High–stakes accountability in texas reconsidered, 2015.
16. Stephen Jay Gould. *The mismeasure of man*. WW Norton & Company, 1996.
17. Julie A Greenberg. Decontructing binary race and sex categories: A comparison of the multiracial and transgendered experience. *San Diego L. Rev.*, 39:917, 2002.
18. Yuval Noah Harari. *Sapiens: A brief history of humankind*. Random House, 2014.
19. Alfie Kohn. *The case against standardized testing: Raising the scores, ruining the schools*. Heinemann Portsmouth, NH, 2000.
20. Veronica A Martinez-Cantu. *Inequality and School Performance: The Effect of the No Child Left Behind Act on the Texas Assessment of Knowledge and Skills Test*. University of North Texas, 2007.
21. Jorge C Rangel and Carlos M Alcala. Project report: De jure segregation of chicanos in texas schools. *Harv. CR-CLL Rev.*, 7:307, 1972.

22. Andria McAdoo Schur et al. *Implications of HB5 at the high school level*. PhD thesis, 2015.
23. Chloe Latham Sikes. Competing goals in college and career readiness policy: An equity analysis of the 2013 texas house bill 5. *Texas Education Review*, 2018.
24. Kurt T Taube and Kathryn W Linden. State mean sat score as a function of participation rate and other educational and demographic variables. *Applied Measurement in Education*, 2(2):143–159, 1989.
25. Ready Texas. A study of the implementation of hb5 in texas and implications for college readiness. `http://www.idra.org/wp-content/uploads/2018/05/Ready-Texas-A-Study-of-the-Implementation-of-HB5-in-Texas-and-Implications-for-College-Readiness-pdf`, Sep 2018.
26. Angela Valenzuela. *Leaving children behind: How" Texas-style" accountability fails Latino youth*. Suny Press, 2005.
27. Jesse Washington. Some asians' college strategy: Don't check 'asian,'. *The Associated Press, December*, 3:1011, 2011.
28. Hadley Wickham. Tidy data. *Journal of Statistical Software*, 59(10), 2014.