High School Standardized Testing Participation in the Test-Optional Post-Pandemic Age



Audrey Liang '24
Data Science Major Capstone

Background & Research Question

In the wake of the COVID-19 pandemic in Spring 2020, there have been drastic changes in the college admissions landscape. This has been reflected especially in standardized testing for college admissions, with the number of test-optional post-secondary institutions soaring from 713 to 1,350 to adjust for the unique circumstances that made it difficult for students to take these tests. [1] Fast forward to 2023, the number of undergraduate institutions adopting a test-optional policy has increased to 1,900 institutions, which represents 83% of four year U.S. post-secondary institutions. [2]

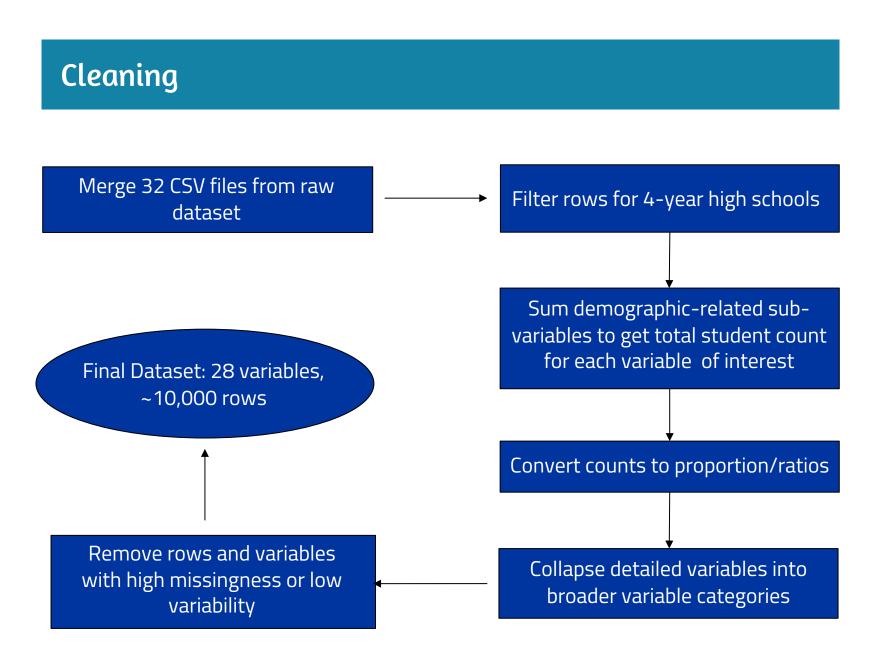
This study uses four-year high school and student body statistics collected from the United States to fit multiple-linear regression model with the research question outlined below:

(RQ) What school and student body characteristics lead to higher standardized testing participation rates, especially in the test-optional age of the pandemic?

Data

The data was collected from the Civil Rights Data Collection Office for Civil Rights [3], which comes from the U.S. Department of Education. This dataset includes information relevant to students' equal access to educational opportunities in the school year of 2020–2021 and was collected from all public schools in the 50 states, Washington D.C., and Puerto Rico.

The final dataset used for this analysis is a merged dataset of various csv files provided from the Civil Rights Data Collection Office for Civil Rights that contain various variables of interest related to school and student characteristics, with each row representing a different public school in the U.S.



Data Exploration/Visualization

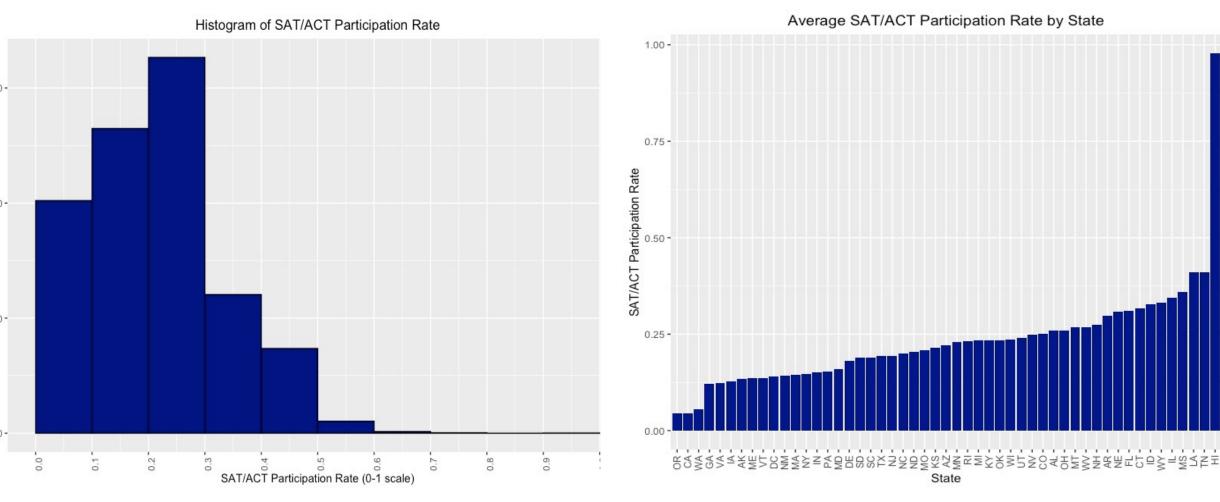


Figure I. Histogram of Standardized Testing Participation Rates Among 4-Year Public High Schools

Figure II. Histogram of Standardized Testing Participation Rates By State (ordered from low to high)

- The graph depicted in Figure I skews right because:
- Most students don't take standardized testing in their first two years of high school
- o Many colleges became test-optional in the 2020-2021 school year and beyond due to constraints of the pandemic
- Taking a closer look at the outliers on both ends of the spectrum in Figure II:
 - o In states like Oregon, California, and Washington that have standardized testing participation rates of ~0.05, most of the public post-secondary institutions in the state became test-optional and still remain as such, not even needing scores for scholarships and other benefits. Many of these schools released statements explaining that standardized testing is not able to predict student success at their institutions. [4] [5] [6]
 - States like Louisiana, Tennessee, and Hawaii have standardized testing participation rates of >0.35. In Louisiana and Tennessee, state law mandates that high-schoolers must take a college entrance exam whether they are attending college or not. Most of the public colleges in-state became test-optional during the pandemic (but still needed scores for scholarships), though many have gone back to mandating reports of standardized testing scores. [7] [8] [9]

First-order Model

A multiple linear regression was run to identify the school and student body related characteristics associated with standardized testing participation rate. Stepwise elimination was performed based on a threshold of 10 for the variance inflation factor to check for multicollinearity. Automatic selection with AIC and BIC criterion was used to identify the best model. The adjusted R-squared for each model were 0.4533 and 0.451 respectively, while the result of 10-fold cross-validation also yielded very similar results. The more parsimonious BIC model with 12 variables was chosen:

Test Participation Rate = AdvancedMath + ApEnroll + Algebra II + Biology + Calculus + Chemistry + DualEnrollment + EslProgram + Disabilities + Referrals + Sports + State

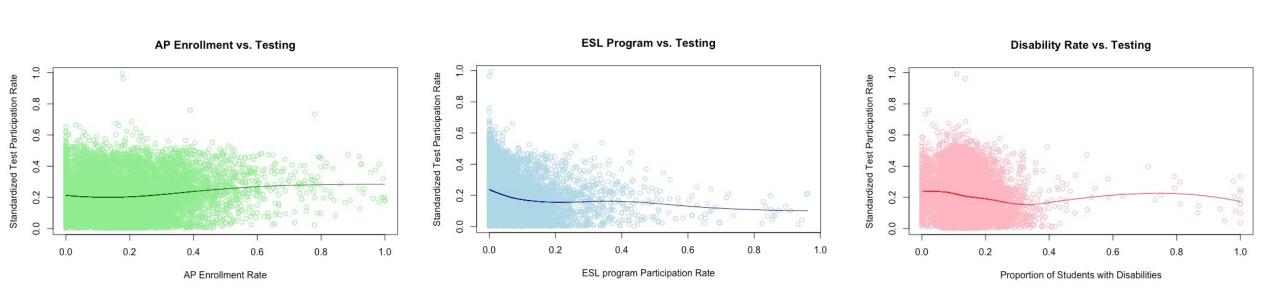


Figure III. Relationships between select predictors (AP enrollment rate, ESL program rate, and proportion of students with disabilities) plotted against standardized test participation rate

Higher-order Model

Interaction terms were added to the first-order model using a AIC and BIC stepwise process. The AIC model yielded a higher adjusted R-squared score of 0.4938, while the BIC model yielded a 0.4549 adjusted R-squared. However, the AIC model had 34 predictors while the BIC model had 20 predictors. To preserve interpretability, the parsimonious BIC model with 20 predictors and a slightly higher adjusted R-squared was chosen as the best higher-order model:

Test Participation Rate = AdvancedMath + ApEnroll + Algebra II + Biology + Calculus + Chemistry + DualEnrollment + EslProgram + Disabilities + Referrals + Sports + State + Calculus*Chemistry + ApEnroll*Sports + AdvancedMath*Disabilities + Calculus*EslProgram + Calculus*:Disabilities + ApEnroll*Calculus + Referrals*Sports + Calculus*DualEnrollment

To choose between the first-order and higher-order model, a partial F-test was conducted to compare the two models. The partial F-test was <0.05, which indicated that the interaction terms in the higher-order model significantly improved the fit of the simple nested first-order model. Thus, the higher-order model is used.

Upon checking for assumptions of linear regression, it was found that the equal variance and linearity assumptions were violated. To solve this problem, a box-cox transformation was performed, which led to the response variable being square rooted. Additionally, all influential outliers were removed from the dataset. Below is a subset of statistically significant output from the model (adjusted R-squared 0.61):

P-value < 2e-16 < 2e-16 < 2e-16
< 2e-16
< 2e-16
0.04.40
3.21e-12
< 2e-16
< 2e-16
< 2e-16
1.71e-08
4.25e-08
2.36e-13
3.63e-14
< 2e-16

Figure IV. Output of final model. The model was run with Alaska as the baseline state (reflected in the intercept). Advanced Math, AP Enrollment, Algebra II, Biology, Dual Enrollment, ESL Program, and Sports refers to proportion of students participating in these classes or programs. Disability refers to the proportion of students with a disability, and referrals refers to students who have been referred to law enforcement for violation of discipline.

Results & Discussion

- Participation rates **vary greatly across states** due to different state and in-state institution policies that this study does not examine in-depth
- Schools where more students take **high-level classes** and participate **in extra-curricular activities** such as pre-college courses and sports tend to have a **higher rate of standardized test participation**.
- Schools where a bigger proportion of students are **not native language speakers or have disabilities** face barriers that might **prevent them from taking standardized tests**, possibly exemplified due to the pandemic.
- Unexpected coefficient estimates are for Biology and Referrals. One would expect a positive association between higher biology course participation rate and test rate and a negative association between higher referral to law enforcement and test rate. These areas are a point for further study.
- Even though most post-secondary institutions across the country are test-optional now, there are still many institutions that require test scores generally, or to be considered for scholarships and programs. It is important to increase accessibility of standardized test preparation and the tests themselves for disadvantaged students in order to increase their pool of choices.
- It is Important to remember that this data is from 2020-2021 when the pandemic and test-optionality first started being implemented on a large-scaled, therefore these insights not necessarily generalizable to other time frames. A future direction study could be a longitudinal study examining the impact of the testoptionality policy change in institutions.