

An Evaluation of the Factors Influencing College Graduation Rate  
June 23, 2020

**Abstract**

Since many people think of a college degree as one of the essential steps on the road to success, we explored the factors that may be associated with the 6-year graduation rate of U.S. colleges and universities that grant at least a bachelor's degree and are not majority online. Our findings are: (a) there is no significant difference among graduation rates in women-only, men-only, and co-ed colleges, (b) there are significant differences in graduation rates among White, Asian, Hispanic, and Black students, (c) a strong correlation exists between average SAT score and graduation rate, and (d) our regression model including household income, average SAT score, and percentage of pell grant students explains over half of the variance in graduation rate. We conducted a cross-sectional observational study which showed that race, household income, average SAT score, and percentage of Pell grant students may be related to a school's graduation rate, but whether a school is coeducational is not.

## Introduction

According to a study done by the Economic Mobility Project, the “key to today’s economy is a college degree,” and yet, out of 100 high school students, only about 30 will make it to college graduation (Beach, 20-23). Since a college degree can be such an important factor in economic mobility and later success, we investigated factors that may influence college graduation rate. We focused on the enrolled students’ scores on the Scholastic Aptitude Test (SAT), students’ median household income, the percentage of students that receive Pell grants, the race of enrolled students, and the type of school (men only, women only, or co-ed).<sup>1</sup>

We chose these factors partially as a result of previous studies that we’ve seen investigate graduation rates. Dougherty et al (2006) explored the relationship between graduation rate and participation/success on the Collegeboard Advanced Placement (AP) exams, very similar to the SAT. They found that high schools with a high percentage of students that took and passed AP exams were better at preparing their students to graduate from college (Dougherty, 3). Seeing the similarities between AP exams and SAT, we wonder if the SAT has a comparable relationship to graduation rate. Additionally, in Sara Hebel’s article (2007), she wrote that “students from low-income backgrounds are less likely to graduate from college than their wealthier peers.” Therefore, we decided to use the percentage of students who are Pell grant recipients as a predictor of graduation rates, in addition to simple household income. We hypothesize that both demographic (race, gender, and finances) and academic factors (SAT score) are related to the graduation rate of an institution.

## Methods

We obtained data from the Department of Education’s College Scorecard, the “Most Recent Institutional Data” file, last updated on March 30, 2020. The 2017-2018 school year is the most recent year reflected in the data. We edited down the original dataset from 6806 to 1276 colleges by choosing to focus our analysis solely on schools that award bachelor or graduate degrees and are not online-only.

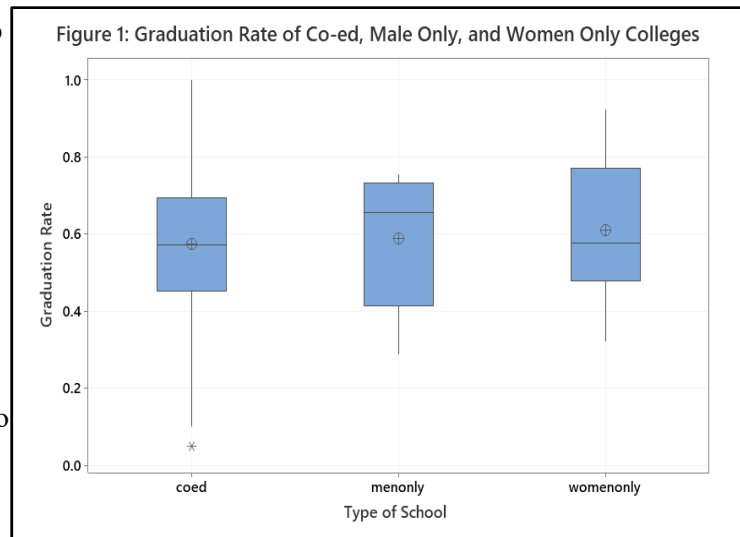
For this project, we used the one-way ANOVA test to explore differences in 6-year graduation rate based on the categorical variables race and type of college (co-ed, women only, or men only). For the race variable, we chose to only focus on Asian, Black, Hispanic, and White students because they are the four most common racial demographics in the United States and in our colleges data set. For the co-ed testing, each school provided an average graduation rate and we compared them across different types of schools. For the race test, we used four graduation rates given by the schools, one per racial category. Not every school reported all four racial graduation rates, but we had at least 1137 cases for each racial category. We took the Asian, Black, Hispanic, and White graduation rates that each institution provided and averaged them to get a mean graduation rate for each race. Then we compared them using a one-way ANOVA test. The residuals of both ANOVA tests are normally distributed, so both tests satisfy that assumption and we feel comfortable making conclusions from these tests. Next, we used Pearson’s correlation coefficient to measure a relationship between two quantitative variables: average SAT score and graduation rate. Then, we used a multiple regression model to show the relationship of average SAT score, median household income, and percentage of students with Pell grants to graduation rate. We converted the original graduation rate numbers from decimals to percentages when we created the multiple regression model due to the large dollar values for the income variable. To carry out the statistical tests and to format our data, we used the statistical software Minitab 19.

---

<sup>1</sup> Pell grants are scholarships given by the federal government to students with outstanding financial need to help them earn a degree (Federal Student Aid).

## Results

First, we looked at the relationship between the types of the college and their graduation rate. The three categories that we focused on were: women-only schools, men-only schools, and coed schools. The residuals for this test were normally distributed, satisfying our assumptions. From our ANOVA test, we received an F-statistic of 0.57 on 1267 df and a p-value of 0.563. This p-value was significantly above the significance threshold of 0.05, so we concluded that institutional graduation rates hardly vary across colleges with different student bodies. Figure 1



corroborates these findings; visually there does not seem to be a meaningful difference in the means across categories.

Next, we used ANOVA testing again to determine if there was a difference in graduation rates of students of different races, exploring 4 races as explanatory categories for this test: White, Black, Hispanic, and Asian. The test satisfied our assumptions, having normally distributed residuals. This test had an F-statistic of 126.91 on 4859 df and resulted in a p-value of approximately 0. The result showed strong evidence that the mean graduation rate varies widely across races. Our Tukey's HSD test revealed that the adjusted p-values for the race pairwise comparisons were all less than 0.05 (approximately 0.000), except for the White & Asian pairing ( $p = 0.937$ ). The mean graduation rates of Black and Hispanic students are lower than the other two race groups. We are 95% confident that the true mean graduation rate of Black students is between 43.9% and 46.4% and the true mean graduation rate for Hispanic students is between 51.1% and 53.6%. Both intervals fall below the equivalent confidence intervals for Asian and White students. Figure 2 supports our conclusions, showing a significant difference across the means of all categories except White and Asian.

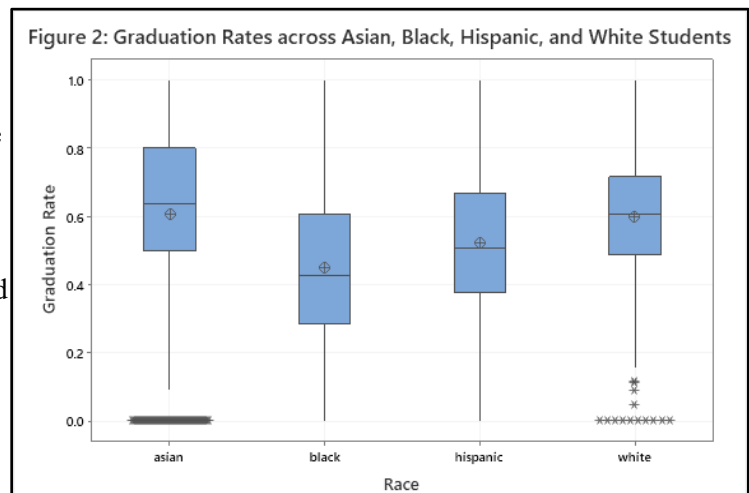


Figure 3: Graduation Rate Regression Model

$$\text{PCTGRAD\_RATE} = -64.77 + 0.08754 \text{ SAT\_AVG} - 10.41 \text{ PCTPELL} + 0.000406 \text{ MEDIAN\_HH\_INC}$$

Third, we looked at the relationship between the average SAT score of students at the institution and graduation rate of the institution by testing whether there was a correlation between them. We got a Pearson's correlation coefficient of 0.811 and a 95% confidence interval of (0.791, 0.829). This strong correlation allows us to conclude that higher SAT scores are very likely to be associated with high

graduation rates. Finally, our regression model in Figure 3 has an adjusted  $R^2$  value of 71.65%, showing that the variables in the model account for 71.65% of the variance in the graduation rate variable. In this model, a 100 point increase in average SAT score corresponds with a 8.754% increase in graduation rate, holding other variables constant (fig. 3). The equation also shows that a single percentage point increase in the percentage of Pell grant recipients is associated with a 10.41% decrease in graduation rate; for a \$10,000 dollar increase in median household income, there is a 4.06% increase in graduation rate.

## Discussion

We found that there is no significant difference among graduation rates in women-only, men-only, and co-ed colleges, which did not support our original hypothesis. The category of whether a college is co-educational is not one of the factors that may influence their graduation rate. Oppositely, we found significant differences in graduation rates between the four race groups that we focused on: White, Asian, Hispanic, and Black. Therefore, consistent with our hypothesis, race is one factor that may affect the college graduation rate. Based on the large correlation coefficient ( $r = 0.811$ ) between average SAT score and graduation rate, we found that high SAT scores are associated with high graduation rates. Our regression model was able to predict 71.65% of the variance in the graduation rate and showed that income and percentage of students with Pell grants are both statistically significant predictors of graduation rate ( $p \approx 0$ ).

For our ANOVA test of the graduation rates for types of colleges, some of our sample sizes were small. We had only 5 men-only colleges and 28 women-only colleges, which could have affected our test. Despite the sample sizes, the p-value of 0.563 is decisively large, so we still conclude that the gender of the student body is unrelated to graduation rate. Secondly, some colleges either did not report their average SAT score or have SAT score-optional policies, which may affect the accuracy of the correlation coefficient and regression model. Even without the non-reporting schools though, we are confident that the large sample size ( $n = 1276$ ) allowed us to form accurate conclusions. When we performed the ANOVA test of race and graduation rate, we also noticed that some colleges did not report their graduation rate by race, which may have affected our results. While there were some reporting inconsistencies for graduation rate by race, in the end, each race group had between 1,100 and 1,260 samples, so we feel confident about the conclusions we drew from the ANOVA. We also acknowledge that there may be a correlation between the graduation rates of different races for a given college, and those correlations further limit our analysis. For example, because graduation rates at an Ivy League school are high for all races, we could accurately predict a graduation rate for one race knowing that it is correlated with another race's graduation rate. In future research, we suggest using a repeated measures analysis to address this limitation. Lastly, to address the possibility of Type I errors in our ANOVA test for race, we used Tukey's HSD test to calculate adjusted p-values for multiple hypothesis tests.

Due to the nature of an observational study, we were not able to conclude any causation relationships from the study, only associations. In theory, it would be interesting to test some of these factors in controlled experiments, but future experimental research may face difficulty due to the impossibility of randomly assigning biological variables like race and gender. Furthermore, we predict that many of these explanatory variables are related to one another, and their relationships could be explored in the future as well (for instance, the relationship between race and income). These relationships are important not only for colleges' self-improvement, but also for equity programs enacted by institutions such as governments, churches, and nonprofits. Finally, we hope our findings can help guide students who are concerned about graduation rates as they make their personal college decisions.

## Works Cited

- Beach, William, et al. *Pathways to Economic Mobility: Key Indicators*. Economic Mobility Project, An Initiative of The Pew Charitable Trust, 2008. Web. 2 May 2020. <[www.pewtrusts.org/-/media/legacy/uploadedfiles/wwwpewtrustsorg/reports/economic\\_mobility/pewempchartbook12pdf.pdf](http://www.pewtrusts.org/-/media/legacy/uploadedfiles/wwwpewtrustsorg/reports/economic_mobility/pewempchartbook12pdf.pdf)>.
- Dougherty, Chrys, et al. *The Relationship between Advanced Placement and College Graduation*. Report 1. National Center for Educational Accountability, 2006. [files.eric.ed.gov/fulltext/ED519365.pdf](http://files.eric.ed.gov/fulltext/ED519365.pdf). Accessed 2 May 2020.
- Federal Student Aid. "Federal Pell Grants." Federal Student Aid, U.S. Department of Education, [studentaid.gov/understand-aid/types/grants/pell](http://studentaid.gov/understand-aid/types/grants/pell). Accessed 5 May 2020.
- Hebel, Sara. "The Graduation Gap. (economically Disadvantaged Students)." *The Chronicle of Higher Education* 53.29 (2007): 1. Print.
- Stumpf, Heinrich, and Julian C. Stanley. "Group Data on High School Grade Point Averages and Scores on Academic Aptitude Tests As Predictors of Institutional Graduation Rates." *Educational and Psychological Measurement*, vol. 62, no. 6, Sage Publications, Dec. 2002, pp. 1042–52, doi:10.1177/0013164402238091.
- U.S. Department of Education, College Scorecard. *Most Recent Institution-Level Data*. (148 MB CSV). U.S. Department of Education, 2020. Web. 23 April 2020. <[collegescorecard.ed.gov/data/](http://collegescorecard.ed.gov/data/)>.