Which Factors Impact Pell Grant Students' Persistence and Graduation?

Faxian Yang
University of South Carolina

Yangzi Mao
Stevens Institute of Technology

Pell Grants are awarded to students from low-income families. The federal government has spent significant resources to support this demographic of students with the goal of improving their socioeconomic status by increasing their likelihood of academic success. The outcome of these efforts has drawn nationwide attention. Key indicators of academic success are retention and graduation rates. This study looked beyond these indicators towards other factors that might had influenced Pell recipients' persistence along with factors that also influence the path to graduation. This study focused on freshmen enrolled in the fall of 2012 at a research university who were awarded Pell Grants. Logistic regression was used in the selection of the significant influential factors separately. The factors, which included GPA by the end of the first academic year, the number of STEM and English courses taken by the end of the second academic year, as well as whether the students successfully passed the courses, etc. were then individually used to predict students' persistence and graduation status. The study results not only indicated that Pell recipients do face more challenges to graduate, but also DID emphasize the critical importance of STEM and English courses for all students.

Keywords: pell grant, persistence/graduation rate, first-generation

INTRODUCTION

The present study focused on freshmen enrolled in the fall of 2012 at a research university. The six-year graduation rate of the overall cohort at the university was 77%, and the corresponding rate for Pell Grant recipients was 69%. The Pell Grant recipients were a subgroup of the overall cohort. The two main research objectives of the present study were to understand and elucidate factors that influence retention and graduation rates of Pell Grant recipients. Furthermore, this study sought to ascertain whether the Pell Grant recipients faced more challenges compared to their fellow students who did not receive Pell Grants. Logistic regression was used to select the significant influential factors in the study separately. Then, these selected factors were employed to predict students' persistence or graduation status individually. The significance level for entry into the two models was 0.05. All selected variables must meet the criteria. The dependent variable was whether the students persisted to the fall of the third year or were awarded a bachelor's degree within six years. The independent variables of the two models included five categories: the students' demographic information, socioeconomic status, college preparation, major, and academic performance.

The first analysis was related to factors which impacted the student graduation within the six-year period. Based on the results of the analysis, factors such as students' persistence in the fall of the third year, GPA by the end of the first academic year, as well as the number of English and STEM courses taken were most significant in order of significance to graduation success. Upon further study of the main variable, students' persistence to the fall of the third year, factors like persistence to the fall of the second year, GPA by the end of the first year, and the number of English and STEM courses taken were found to be the most significant in order of significance. According to the analytics above, student graduation was mainly dependent on persistence to the fall of the third year while persistence to the fall of the third year relied on persistence to the fall of the second year. Therefore, persistence became the most significant factor for the students' success, albeit other aforementioned factors also played contributing roles.

What are these other factors that also impact the Pell Grant recipients' academic success? A study by Millea, Wills, Elder, and Molina (2018) indicated that students' academic preparation was highly correlated to their graduation and persistence. In their study, academic preparation was based on students' standardized test scores and high school GPA. Although the fitted model of this study did not use standardized test scores, a count of the six identified STEM courses a student needed to retake were selected instead. This variable fully reflected the students' college preparation in STEM fields since these STEM courses were all introductory level courses in each field, and the seven credits of these science courses fulfilled the requirement of general education. Although the variable for academic preparation was represented differently in Millea et al. (2018) and this study, one with standardized test scores and one with STEM courses retaken, respectively, the results and conclusions drawn from both studies were in concordance. With respect to demographic information, the model generated during the present study did not find that being in an underrepresented minority group to be a factor on graduation rate. Although credits earned from the first-year seminar (UNIV-101) were not selected by the model as a factor affecting graduation rate, it was selected in the model to predict the students' retention in the fall of the third year. Therefore, the first-year seminar had an indirect influence on the students' graduation. Since any majors resulting in a bachelor's degree awarded were counted as academic success, each students' major was not a factor that impacted graduation. This study results emphasized the importance of students' first two years of academic study, especially their learning experiences with English and STEM courses, since in turn, this impacted the students' retention rates in the fall of the second and third years. What was also notable was that the Pell Grant recipients DID face more challenges regarding persistence and graduation success.

BACKGROUND

The literature review included the following three parts: the original reason behind initiating the Pell Grant program and the results of funding these students, the latest update on retention rate and graduation rate in the country and factors which affected these rates and, lastly, statistical methods applied to this study.

Pell Grant

Pell Grants are awarded to students from low income families. Although most Pell Grants recipients come from families with an annual income of less than $20,000, students from families with an annual income up to $50,000 may also be eligible (Federal Pell Grants, 2019). Since creating the Pell Grant program in 1972, the U.S. federal government has contributed significant funding for the program. In 2007-2008, recipients totaled 5.5 million students with an annual program expenditure of $14.7 billion. Then, these numbers further increased in 2011-2012 with recipients totaling 9.7 million and an annual program expenditure of $33.4 billion (Mullin, 2013). With program expenditures topping hundreds of billions of dollars per decade, the efficacy of the program, demonstrated by the rate of successful academic outcomes for Pell Grant recipients, understandably draws much interest from politicians, educators, and economists.

Factors Impacting the Retention Rate and Graduation Rate

Students' retention and graduation rates are key indicators of academic performance and success. For all first-time, full-time, degree-seeking undergraduate students who enrolled in four-year degree-granting

institutions in Fall 2017, the overall second-year retention rate was 81 percent (U.S. Department of Education, 2020). Based on the latest information from the National Center for Education Statistics (U.S. Department of Education, 2020), nationwide, the six-year graduation rate at four-year institutions for the 2012 freshmen cohort was 62%. The corresponding rate was 61% at state institutions, 67% at private nonprofit institutions, and 25% at private for-profit institutions (U.S. Department of Education, 2020).

Huang, Roche, Kennedy, and Brocato (2017) observed that most research classified factors that affected graduation success rate into three categories: demographic, pre-college, and sociological factors. In addition to these factors, the authors also investigated how factors such as the student's college major, home address, and use of learning support affected a student's likelihood of graduating within six years. Chatterjee, Marachi, Natekar, Rai, and Yeung (2018) identified that both total credits accumulated at the end of the first year, and first-year retention rate have significant positive effects on graduation success. In addition to selecting Pell Grant recipients, first-generation students were also used as another important sociological factor indicator in the prediction of graduation rate (Bird, 2018). For the pre-college factor, English composition and introductory courses of STEM subjects were used in the model since those courses have more direct connections to high school coursework and college preparation.

Logistic Regression and Stepwise Regression

Logistic regression can be binomial, ordinal, or multinomial. The first, binomial logistic regression, is best situated for when there can only be two possible observed outcomes for a dependent variable, "0" and "1". For example, "0" may represent "not graduated" while "1" may present "graduated" (Logistic regression, n. d.). The binary indicators of graduation success and persistence in the third fall, where their above median or below median graduation/persistence in the third fall, along with hypothesized covariates requires the utilization of logistic regression to facilitate the interpretation of the effects by utilizing odds ratios and probabilities (Morrison, 2012). "Odds ratio compares the odds of the yes proportion for the Group 1 to the odds of the yes proportion for the Group 2" (Stokes, Davis, & Koch, 1995, p. 28). This study employed odds ratio to compare the graduation rates of the students receiving Pell Grant to the students who did not.

"Stepwise regression provides an important modification of forward selection in that, at each stage of selection, all repressors currently in the model are evaluated through the partial F-test. A preselected F_{out} critical value is used" (Myers, 1989, p. 187).

METHODOLOGY

Participants

The participants were 4,578 students in the 2012 adjusted cohort at a research university. Based on the original cohort, the term, adjusted cohort, is meant for calculating retention and the graduation rates after removing any allowable exclusions, like death, permanent disability, service in the armed forces, service with a foreign aid service of the federal government, or service on official church missions (IPEDS, 2020). Of the students in the adjusted cohort, 862 students were awarded Pell Grants. Since some students did not have all the necessary independent variables used to predict their graduation, such as standardized test scores, these students were excluded. The analytical models only included the students who possessed all variables in the model. Of the 4,578 students, 3,251 students were used in the model. Of the students in the model, 613 students were Pell Grant recipients, and 2,638 were non-Pell Grant recipients.

Dependent Variable and Independent Variables Considered

The dependent variables were whether the 3,251 students had been awarded a bachelor's degree within six years and if they persisted on to their third year successfully. Based on previous studies done by Huang et al. (2017) and Chatterjee et al. (2018), the independent variables spanned five separate categories: demographic related information, socioeconomic status, college preparedness, college major, and academic performance.

Demographic Related Information

Demographic information of interest in this study included SEX (male or female) and RACE (underrepresented minority group or not). If a student's race was neither white nor Asian, the student was coded within the underrepresented minorities group. Otherwise, the student was not categorized to be in the underrepresented minorities group.

Socioeconomic Status

The status of a student as a first-generation college student (_1ST_GEN) was coded as one, otherwise coded as zero. The status of a student as a recipient of Pell Grant (PELL) was coded as one, otherwise coded as zero.

College Preparation

College Board (2018) had indicated that the multiple correlation of SAT scores and high school GPA with first-year college GPA continues to be strong for the 2013 cohort (r = .61). Since the composite SAT score is an important indicator of first-year college GPA, the sum of the SAT (2005 Edition) scores for reading and math was used for this study. If students had submitted ACT scores in lieu of SAT scores, their ACT scores were converted to SAT scores based on the concordance table provided by the College Board.

The reason that the variable of students repeatedly taking introductory STEM courses was selected in the model was to test whether the students had enough college preparation, specifically in STEM fields. All six STEM courses were introductory level courses of each field, and all students were required to complete seven credits in science subjects and six credits in math or statistics or computer science subjects in order to complete any given bachelor's degree at this university.

College Major

Since students in STEM majors faced more rigorous academic challenges such as increased math and sciences course prerequisites than non-STEM majors, did these challenges impact their graduation rates? Similarly, did these challenges force STEM majors' students to change to non-STEM majors in the third fall?

Academic Performance

Eight variables were related to students' academic performance. They were the number of English courses taken (ENGL_COUNT), the number of math courses taken (MATH_COUNT), first-year GPA (GPA_1ST_YEAR), final grade earned in the English course (ENGL_POINT), final grade earned in the math course (MATH_POINT), retention in the third fall (BACK_3RD_FALL), credit hours earned in the first year (EARN_HOURS_1), and course credits earned from freshmen seminar (UNIV-101). Points of UNIV-101 were converted by course credits earned from freshmen seminar. A value of "-1" was assigned when a student chose not to take the seminar.

Analytical Methods

Multicollinearity

Multicollinearity is a common problem when estimating linear or generalized linear models, and logistic regression is no exception. This occurs when there is high correlation among independent variables. When the study was initially designed, there were 15 variables in the model. To combat the multicollinearity problem, UNIV-101, EARN_HOURS_1, BACK_3RD_FALL and MAJOR_CHANGE_3RD_FALL were removed because they were so highly correlated to cumulative GPA at the end of the first academic year.

Independent Variables Used

The 11 independent variables mentioned above were employed to predict students' graduation status. They were Pell Grant recipient or not, first-generation college student or not, number of English and math courses taken separately, first-year GPA, final grades earned in both English and math, sex, underrepresented minority group or not, SAT score, and number of STEM courses retaken. The logistic

regression was used to test whether graduation was jointly independent of the independent variables simultaneously. H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = 0$. All selected variables must meet the criteria. The formula above can be interpreted as the hypothesis that the 11 independent variables did not have any impact on students' graduation success.

RESULTS

Stepwise Logistic Regression Model

Since 3,251 students were used in the model, their descriptive statistics of four college preparation and academic performance related variables were presented. Average GPA after the first year was 3.288, average counts of English and math courses taken were 2.235 and 2.453, respectively. Since 50.38% of the students did not retake any of the six STEM courses, the average count of retaking the six STEM courses was 1.061. Although the average count diluted the importance of the variable, retaking any of the six STEM courses was, in fact, a significant variable which predicted the students' graduation.

In the fitted model, of the six courses listed in Table 1, two were introductory level courses for both chemistry and biology. Calculus I was the first course in the three series of calculus. Because physics was too challenging for students, and thus, very few students selected courses in this field, physics was not selected. One thousand six hundred thirty-eight students (50.38%) did not retake the six STEM courses. The remaining students retook at least one of the courses. 691 students (21.25%) retook one of the courses, 341 (10.49%) retook two, 319 (9.81%) retook three, 192 (5.91%) retook four, and 70 (2.15%) retook five.

TABLE 1
REPEAT ENROLLMENT OF SIX STEM COURSES

Course	Course Title	Sum of Enrollment Time of Each Course	Enrollment Time							
			2	3	4	5	6	7	8	9
CHEM-111	General Chemistry	1,314	1,210	30	67	6	1	0	0	0
BIOL-101	Biological Principles I	967	785	152	26	3	0	1	0	0
CHEM-112	General Chemistry	708	28	633	32	5	10	0	0	0
MATH-141	Calculus I	640	9	500	46	5	70	6	0	4
GEOL-101	Intro to the Earth	597	597	0	0	0	0	0	0	0
BIOL-102	Biological Principles II	546	505	36	3	2	0	0	0	0
Summary		4,772	3,134	1,351	174	21	81	7	0	4

As indicated in Table 2, GPA at the end of the first year, the number of mathematics courses taken, and the number of English courses taken had positive effects on graduation. Therefore, the larger the values of the coefficients of these three variables, the higher the probability of graduation success. However, a negative coefficient was observed for each of the three following variables: count of six STEM courses a student needed to retake, whether or not the student was the recipient of a Pell Grant, and whether or not the student was a first-generation college student. This meant that these three variables had negative effects on graduation. The smaller the value of the three coefficients, the less the probability of graduation success.

TABLE 2
ANALYSIS OF MAXIMUM LIKELIHOOD ESTIMATES FOR PELL GRANT RECIPIENTS

Parameter	DF	Estimate (Coefficients)	Standard Error	Wald Chi-Square	Pr>ChiSq
Intercept	1	-3.324	0.2953	126.737	<.0001
GPA_1ST_YEAR	1	1.0661	0.0787	183.6004	<.0001
MATH_COUNT	1	0.4452	0.0469	89.9318	<.0001
ENGL_COUNT	1	0.2576	0.0461	31.2751	<.0001
6_STEM_COURSE	1	-0.1014	0.0354	8.1843	0.0042
PELL	1	-0.2738	0.1188	5.3173	0.0211
_1ST_GEN	1	-0.3659	0.1416	6.6729	0.0098

The six independent variables in Table 2 were equally applicable to the prediction of graduation rates of both Pell Grant recipients and non-Pell Grant recipients. When the rate of Pell Grant recipients needed to be predicted, PELL variable equaled one. However, when the rate of non-Pell grant recipients needed to be predicted, PELL equaled to zero.

The three variables, first-year college GPA, first-generation student, and Pell Grant recipient confirmed the long-held assumption of education professionals nationwide that total credits accumulated at the end of the first year has significant positive effects on graduation success (Chatterjee et al., 2018). Although first-generation students made up an increasingly large segment of the high-school population, they were still underrepresented in post-secondary education. They were very much underrepresented in four-year colleges and universities. Moreover, even if these first-generation students enrolled, they were less likely than their peers to graduate (Conefrey, 2017). "The higher the percentage of Pell recipients, the lower the institutional probability expectations for exceeding median college performance in graduation rates" (Morrison, 2012, p. 168).

The number of math and English courses taken separately were selected, and their impact was highlighted by this study. That both estimated coefficients were positive in the fitted model can be interpreted to mean that the more credits earned from either math or English courses equated to a higher probability for graduating.

After running stepwise logistic regression model in SAS, the fitted graduation model can be written as follows:

$$\text{PROB (GRADUATION =1|(x))} = \lambda\,(\alpha + \beta_{1\text{ENGL\_COUNT}}\,\text{ENGL\_COUNT} + \beta_{2\text{MATH\_COUNT}}\,\text{MATH\_COUNT} + \beta_{3\text{ GPA\_1ST\_YEAR}}\,\text{GPA\_1ST\_YEAR} + \beta_{4\text{ \_1ST\_GEN}}\,\text{\_1ST\_GEN} + \beta_{5\text{ Pell}}\,\text{PELL} + \beta_{6\text{ 6\_STEM\_COURSE}}\,\text{6\_STEM\_COURSE})$$

Hypothesis Tested

The Global Null Hypothesis: $\beta = 0$ was tested. This logit model was statistically significant. The reported likelihood-ratio (LR) tested that GRADUATION was jointly independent of the predictors simultaneously; $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$. The LR test statistic of 401.8295 was chi-squared (χ^2) with 6 degree of freedom and *p*-value <0.0001. In addition, all predictors of GRADUATION were statistically significant as their *p*-values are less than 0.05. Hosmer and Lemeshow Goodness-of-Fit Test also provided additional evidence the model fits the data well ($\chi^2 = 14.3564$; $df = 8$; p=0.0729).

The Fitted Model Applied

For the sake of making the above formula more meaningful, the coefficients in Table 2 were plugged into the formula above. The formula calculation was as following:

PROB (GRADUATION =1|(x)) = λ (-3.324 + 0.2576 ENGL_COUNT + 0.4452 MATH_COUNT + 1.0661 GPA_1ST_YEAR - 0.3659 _1ST_GEN-0.2738 PELL-0.1014 6_STEM_COURSE)

Supposed a Pell Grant recipient took two math and two English courses. This student's first-year GPA was 3.000, and the student retook two of the six STEM courses. The probability of earning a degree within six years was

PROB (GRADUATION =1|(x)) = λ (-3.324 + 0.2576 X 2 + 0.4452 X 2 + 1.0661 X 3 - 0.3659 X 0 - 0.2738 X 1-0.1014 X 2)

= λ (-3.324 + 0.5152 + 0.8904 + 3.1983 - 0 -0.2738 – 0.2028) = λ (0.8033)

$$\pi(x) = \frac{e^{0.8033}}{1+e^{0.8033}} = 69.07\%$$

The calculation above indicated that the Pell Grant recipient had 69.07% chance of graduation in the fitted model, which was as same as the rate observed. However, if this student was also a first-generation college student and holding the other four variables constant, the probability of graduation rate for a Pell Grant recipient and a first-generation college student went down to 60.77%. This calculation was based on the student's first-year GPA of 3.000. If the student's GPA was below 3.000, the chance of graduation would further diminish. The first-year GPA contributed to the largest positive weight on graduation success. For example, a student, who was neither a Pell Grant recipient nor a first-generation college student, took two math and two English courses, did not retake any of the six STEM courses, and had an average first-year GPA of 3.288 would then have a predicted graduation success rate of 83.01%. It was not only higher than the students who had been awarded Pell Grant and were first-generation college students (of which the rate was 60.77%), but it was also higher than Pell Grant recipients who were not first-generation college students (of which the rate was 69.07%).

Odds Ratio Applied

Morrison (2012) indicated that it is not easy to interpret the above statistics, since the coefficients are quite different from ordinary least squares. Because dichotomous probability is a 0 or 1, outcome that is a nonlinear function of the logit, and the logit coefficient indicates how much the logit increases for ONE unit of change in the independent variable, the statistics may be more easily interpreted by converting the logit coefficients into odds ratios.

In contrast with the fitted model discussed earlier, odds ratio focuses on comparing outcomes of one independent variable instead of all variables selected. Therefore, it offers more intuitive and easily interpreted results. Odds ratio equals to $e^{\beta i}$. As discussed previously, six independent variables were significant in predicting graduation rate. In the fitted model above, the coefficient of PELL equaled to -0.2738, or $\beta_{5\ PELL} = -0.2738$. Therefore, its odds ratio equaled to $e^{-0.2738} = 0.76$ in Table 3.

TABLE 3
ODDS RATIO ESTIMATES

Effect	Point Estimate	95% Wald Confidence Limits	
_1ST_GEN	0.6940	0.5250	0.9160
ENGL_COUNT	1.2940	1.1820	1.4160
MATH_COUNT	1.5610	1.4240	1.7110
GPA_1ST_YEAR	2.9040	2.4890	3.3880
6_STEM_COURSE	0.9040	0.8430	0.9690
PELL	0.7600	0.6030	0.9600

Holding all other variables constant, the probability that a Pell Grant recipient would graduate was 24% (1 - 0.7600) less than the probability for a student who did not receive a Pell Grant. Applying the same methodology, $\beta_{4\_1ST\_GEN} = -0.3659$, its odds ratio equaled to $e^{-0.3659} = 0.6940$. Holding all other variables constant, the probability that a first-generation college student would graduate was 31% (1- 0.6940) less than the probability for a student who was not a first-generation college student. The odds ratios were consistent with the coefficients from the fitted model, though they were shown in different ways. Since these groups of students were in comparatively weaker positions while pursuing their degrees in college, students in either of these two groups need to be supported extensively.

Contrast Comparison

"Contrast specifies the type of contrast that is used for categorical independent variables. The interpretation of the regression coefficients for categorical variables depends on the contrasts that are used" (IBM Knowledge Center, n. d.).

The function contrasts a specific independent variable's change either in a positive or negative direction while holding other variables constant. The function presents results of dependent variable's change with the specific independent variable's change. Table 4 presents how GPA by the end of the first year affects graduation rate for Pell and first-generation students while holding other independent variables constant. If a student was a Pell Grant recipient and first-generation, then the student's graduation rate was 87.68% when the first-year GPA was 4.000. The rate was decreased to 80.69% when the GPA was 3.500, etc.

TABLE 4
CONTRAST ESTIMATION AND TESTING RESULTS BY ROW

Contrast	Type	Row	Estimate	Standard Error	Alpha	Confidence Limits		Wald Chi-Square	Pr> ChiSq
GPA_1ST_YEAR =2.500	PROB	1	0.5899	0.0433	0.05	0.5033	0.6713	4.1346	0.0420
GPA_1ST_YEAR =3.000	PROB	1	0.7103	0.0410	0.05	0.6240	0.7836	20.3011	<.0001
GPA_1ST_YEAR =3.500	PROB	1	0.8069	0.0350	0.05	0.7291	0.8664	40.6165	<.0001
GPA_1ST_YEAR =4.000	PROB	1	0.8768	0.0274	0.05	0.8125	0.9212	60.0592	<.0001

Confusion Matrix

The confusion matrix is a method that this study used to compare the observed graduation status (graduated or not) to the graduation status from the fitted model (predicted to graduate or not). As such, the confusion matrix method could be used to identify the two subgroups of students, one who was predicted to graduate and the other who was predicted to not graduate. There were 3,251 students in the model. First, there were 2,543 students, who were predicted to graduate and DID graduate. However, there were also 112 students who were predicted to not graduate and accordingly, DID NOT graduate. Therefore, the accuracy rate of this model was $\frac{2543+112}{3251} = 81.67\%$. The 540 students whom the model predicted would graduate but did not actually graduate were classified as the "at-risk" group. The 56 students whom the model predicted would not graduate but in fact did graduate were classified in the "learning group" (Chatterjee et al., 2018). The ability to identify the students, who might not graduate, could potentially greatly assist academic departments' and student services' abilities to locate and assist this group of students when they can still turn things around.

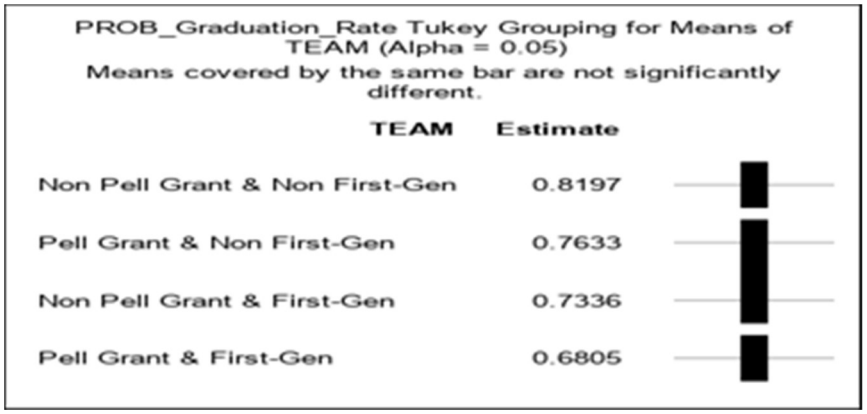
ANOVA

The ANOVA analysis served to double check the logistic regression results presented earlier. Either Pell Grant recipients or first-generation college students were less likely to complete their degrees. The 3,251 students in the fitted model had been categorized into four groups based on their Pell Grant recipients and first-generation statuses. These groups were Pell Grant and first-generation college students, Pell Grant and non-first-generation college students, non-Pell Grant and first-generation college students, and non-Pell Grant and non-first-generation college students. Based on the ANOVA analysis results, overall graduation rates of the students in the four groups were significantly different ($F = 84.82$, $\alpha = <.0001$).

As seen in Figure 1, the graduation rate (PROB_Graduation_Rate) of Non-Pell Grant & Non-First-Gen group was 81.97%, and the rate of Pell Grant & First Gen was 68.05%. Both rates were like rates in the Fitted Model Applied section, of which the rate of Non-Pell Grant & Non-First-Gen group was 83.01%, and the rate of Pell Grant & First Gen was 60.77%. The two rates were significantly different, which were shown in different bars in Figure 1. However, the rates of Pell Grant & Non-First Gen and Non-Pell Grant & First-Gen were not significantly different, since each group included students who faced more challenges to graduation (shown in Figure 1). The negative impacts of being either a Pell Grant recipient or a First-Gen student were already explained in the Odd Ratio section of this paper. The probability that a Pell Grant recipient would graduate was 24% less than the probability for a student who did not receive a Pell Grant. The probability that a first-generation college student would graduate was 31% less than the probability for a student who is not a first-generation college student.

The only purpose of employing ANOVA was to cross check the results from logistic regression since the latter was the major tool of analysis in this study. The results of ANOVA verified the results from logistic regression.

FIGURE 1
TUKEY'S STUDENTIZED RANGE (HSD) TEST FOR GRADUATION RATE



DISCUSSION

Application of This Study

Besides addressing the two research questions of this study, the principal motivation of this study is to apply the results to the real world. As discussed in the Results section, the Confusion Matrix is a straightforward method to achieve this goal since it can predict who may potentially not graduate as early as the end of students' sophomore year. Therefore, this would be the best time to help the students who were predicted not to graduate. In addition, improving students' first-year GPA, increasing the number of math and English courses taken all had positive effects on graduation; on the contrary, the increase in the count of six STEM courses a student needed to retake had negative effects on graduation. These four areas were where the Student Counseling Division may put their efforts. Therefore, at that midpoint of their college career, enhanced support will increase the likelihood that they will go on to eventually graduate.

Academic Performance

Some findings of this study did not come as a surprise to the higher education community at all. Rather, they just confirmed the community's long-held beliefs. For example, both students' persistence and graduation success are highly associated with their first-year academic performance. However, the results of analyzing persistence to the third fall emphasize the importance of English and math courses taken by the end of the second academic year. They were highly correlated with students' academic performance. This may be interpreted as the importance of these two most fundamental subjects. For students with majors in social sciences or humanity, it is better not to overestimate the importance of English courses since reading and writing are foundations of the areas. Similarly, the role of mathematics plays an essential role in STEM majors as logic thinking and many math concepts are crucial skills for students in these fields.

Academic Preparation

A study by Millea et al. (2018) indicated that students' academic preparation was highly correlated to their graduation or persistence. Academic preparation in their study was based on students' standardized test scores and high school GPA. Although the fitted model of this present study did not select standardized scores, count of the six identified STEM courses a student needed to retake was selected in its stead. This replacement variable fully reflected the students' college preparation in STEM fields, since the six STEM courses were introductory level courses of each field. This agreed with the conclusion by Millea et al. (2018), though the variable used for academic preparation differed.

As discussed previously, students' academic performance was a major predictor of graduation and retention. However, their academic performance at the university level reflected their academic preparation and performance at the high school level. If their academic preparation was insufficient, understandably, taking introductory level STEM courses at the college level would be extremely challenging for them. This was clearly shown by the 1,613(49.62%) students who had to retake at least one STEM course to fulfill the requirement of their general education. If the 1,613 students had all passed the courses the first time, the overall graduation rate would have risen significantly. Certainly, while many other factors could also contribute to why these students had to retake these courses, their academic/college preparation at their high schools was surely one of the major factors.

"The percentage of students meeting at least three of the four ACT College Readiness Benchmarks was 37 percent." (ACT, 2019). In other words, 63% ACT takers nationwide in 2019 did not meet ACT College Readiness Benchmarks. The ACT® College Readiness Benchmarks were scores on the ACT subject area tests (English, Math, Reading, and Sciences) that represented the level of achievement required for students to have a 50% chance of obtaining a B or higher or about a 75% chance of obtaining a C or higher in corresponding credit-bearing first-year college courses (ACT, 2019). In comparison, 49.62% students of the study group demonstrated insufficient college preparation. To boost the collegiate graduation and retention rates, college readiness during high school should be heavily emphasized and worked on first. Otherwise, it would be a futile game of catch-up if most of the efforts to better academic performance was during the university years.

Gender and Underrepresented Minority

With respect to demographic information, the model generated during the present study did not find either gender or underrepresented minority status to be factors that influence graduation rate. This agreed with the conclusion by Millea et al. (2018), though other studies which primarily focused on students' majors in science and engineering (NCES, 2019 & NCES, 2020) had different conclusions.

LIMITATION

Students' high school GPAs were not used in the model since some GPAs were weighted, and some were not. For example, a student with a weighted high school GPA of 4.5 cannot be compared to an unweighted high school GPA of 3.9 as the scales are entirely different, and schools may have different GPA weighting scales.

Although standardized test scores and high school GPA have multiple correlation with first-year college GPA (College Board, 2018), standardized scores were not selected for the model. One explanation may be that standardized scores are not as significant as other variables selected, especially the count of repeatedly taken introductory STEM courses, the numbers of English and math courses taken, and first-year college GPA.

Due to considering multicollinearity of this study, the questions of whether the difficulty of the STEM curriculum serves as a blocker for STEM major students towards graduation, or if these students were forced to change to non-STEM majors were not addressed. Both questions would warrant future studies.

CONCLUSION

As discussed above, this model selected six independent variables in the fitted model. The indicator, Pell Grant recipient, was one of the main selected variables. In addition to five other independent variables which impacted all students, the variable of whether a student was a Pell Grant recipient or not greatly affected graduation success. The graduation rate of Pell Grant recipients was 24% less than the rate of their non-Pell Grant counterparts. One hundred ninety-two Pell Grant recipients were also first-generation students. Therefore, if the students were both Pell Grant recipients and first-generation, their graduation rate was even lower than the rate of non-Pell Grant and non-first-generation students. Since this analysis highlighted that Pell Grant recipients were statistically at a greater disadvantage compared to other students in the cohort, there is an established need for the Pell Grant recipients to get more support from various university services to ensure a greater chance at success in persistence and graduation. However, besides the socioeconomic status, colleague preparation and academic performance are also deciding factors for all students to graduate in six years.

REFERENCES

ACT. (2019) *The Condition of College & Career Readiness.* Retrieved from http://www.act.org/content/dam/act/unsecured/documents/cccr-2019/National-CCCR-2019.pdf

Bird, G. (2018). The impact of parents' education levels. *Inside Higher ED*. Retrieved from https://www.insidehighered.com/news/2018/02/08/students-postsecondary-education-arcs-affected-parents-college-backgrounds-study

Chatterjee, A., Marachi, C., Natekar, S., Rai, C., & Yeung, F. (2018). Using logistic regression model to identify student characteristics to tailor graduation initiatives. *College Student Journal*, *52*(3), 352-360.

College Board. (2018). *Validity of the SAT® for predicting first-year grades: 2013 SAT validity sample*. New York: The College Board. Retrieved from https://files.eric.ed.gov/fulltext/ED582459.pdf

Conefrey, T. (2017). LEADing the way with ePortfolios in a first-generation learning community. *International Journal of ePortfolio*, *7*(2), 161-173.

Federal Pell Grants. (2019, February 14). Retrieved from https://studentaid.ed.gov/sa/types/grants-scholarships/pell.

Huang, L., Roche, L., Kennedy, K., & Brocato, M. (2017). Using an integrated persistence model to predict college graduation. *International Journal of Higher Education*, *6*(3), 40-53. Retrieved from https://files.eric.ed.gov/fulltext/EJ1142304.pdf

IBM Knowledge Center. (n.d.) CONTRAST Subcommand (LOGISTIC REGRESSION command). Retrieved from https://www.ibm.com/support/knowledgecenter/SSLVMB_24.0.0/spss/regression/syn_logistic_regression_contrast.html

IPEDS. (2020). Retrieved from https://surveys.nces.ed.gov/ipeds/VisGlossaryPopup.aspx?idlink=207

Logistic regression. (n.d.) In *Wikipedia*. Retrieved from https://en.wikipedia.org/wiki/Logistic_regression#:~:text=Logistic%20regression%20can%20be%20binomial,%22loss%22)

Millea, M., Wills, R., Elder, A., & Molina, D. (2018). What matters in college student success? Determinants of college retention and graduation rates. *Education*, *138*(4), 309-322.
Morrison, M. (2012). Graduation odds and probabilities among baccalaureate colleges and universities. *Journal of College Student Retention*, *14*(2), 157-179.
Mullin, C. (2013). Past, present, and possibilities for the future of the Pell Grant program. *Journal of Education Finance*, *39*(1), 3-14.
Myers, R. (1989). *Classical and modern regression with applications* (2nd ed.). Belmont CA: Duxbury press.
NCES. (2019, February). *Status and trends in the education of racial and ethnic groups.* Retrieved from https://nces.ed.gov/programs/raceindicators/indicator_REG.asp
NCES. (2020, May). *The condition of education-postsecondary education-programs, courses, and completions-undergraduate degree fields*. Retrieved from https://nces.ed.gov/programs/coe/indicator_cta.asp
Scholarships.com. (n.d.). Federal Pell Grants. Retrieved from https://www.scholarships.com/financial-aid/federal-aid/federal-pell-grants/
Stokes, M., Davis, C., & Koch, G. (1995). *SAS/STAT® 13.1 User's guide introduction to categorical data analysis procedures.* Cary, NC. SAS Institute Inc.
U.S. Department of Education. (2000). *Entry and persistence of women and minorities in college science and engineering education.* National Center for Education Statistics. Retrieved from https://files.eric.ed.gov/fulltext/ED566411.pdf
U.S. Department of Education. (2020). *Spring 2020, Undergraduate retention, and graduation rates.* National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS). Retrieved from https://nces.ed.gov/programs/coe/indicator_ctr.asp