BILINGUAL PUBLISHING CO.
Pioneer of Global Academics Since 1984

Journal of International Education and Practice
https://ojs.bilpublishing.com/index.php/jiep

ARTICLE

A Logistic Regression Model to Predict Graduate Student Matriculation

Lei Ouyang[1] Tanjian Liang[2*] Xiuye Xie[3] Sonja Rizzolo[4]

1 School of Kinesiology, Marshall University, Huntington, WV, USA
2 Department of Sport and Movement Studies, Central Washington University, WA, USA
3 Department of Kinesiology, Missouri State University, Springfield, MO, USA
4 Graduate School, University of Northern Colorado, Greeley, CO, USA

ARTICLE INFO

Article history
Received: 28 November 2020
Accepted: 21 June 2021
Published: 30 August 2021

Keywords:
Matriculation
Graduate Admission
Financial Aid
Enrollment

ABSTRACT

Higher education institutions invest a significant amount of resources every year to recruit new students. However, higher education administrators have been continuously facing challenges in enrollment management due to the demographic shifts, dramatic increases in educational costs, intense competition among institutions, and the uncertain nature of human selection patterns (Baum, Kurose, &McPherson, 2013).[3] Today's post-baccalaureate applicants are more knowledgeable than in previous years, because they can access information on a specific graduate program, in a given college, at any time. As reported in numerous studies, the number of graduate students switching out of their universities continues to be an essential issue. A useful prediction model of matriculation that uses available student data is highly desirable to assist the graduate students with timely advising early in their universities. This study was designed to build a predictive model for the probability that a specific admitted graduate student will matriculate. The results indicated that ten predictive variables were statistically significant at the .05 level. Getting an assistantship made the most substantial positive contribution in predicting student matriculation, followed by FAFSA, experience with the university, campus, degree level, college, gender, age, the number of days between application and admission, and distance to the university. This study's results could be beneficial for improving marketing efforts aimed toward individuals with characteristics most likely to enroll. Administrators could calculate the predictive score (or percentage) for each prospective student based on the predictive model. Marketing efforts could then concentrate on those applicants whose predictive score is high and eliminate the low qualifying students from their recruitment plan.

1. Introduction

A Logistic Regression Model to Predict Graduate Student Matriculation

Many universities in the United States (U.S.) have seen an increase in postbaccalaureate degree program enrollment. In fact, between 2000 and 2010, there was a 36% increase in postbaccalaureate enrollment (Snyder, de Brey, & Dillow, 2018).[31] Postbaccalaureate degree programs include master's and doctoral programs, as well

**Corresponding Author:*
Dr. Tanjian Liang, Department of Sport and Movement Studies, College of Education and Professional Studies, Central Washington University, WA, USA; E-mail: Tanjian.Liang@cwu.edu.

DOI: https://doi.org/10.30564/jiep.v4i1&2.2628

as medical, dentistry, and law. According to Snyder et al. (2018),[31] in fall 2015, over 2.9 million students were enrolled in postbaccalaureate degree programs (2016). In addition, these studies estimated that between 2015 and 2026, postbaccalaureate enrollment will increase by 12% (from 2.9 million to 3.3 million students). Graduate education in the U.S. has risen to the top of the international education enterprise (Shanghai Ranking Consultancy, 2009).[30] Worldly university rankings such as *The Times of London* and the Shanghai Jiao Tong University placed the U.S. postbaccalaureate degree programs and their research facilities among the best in the world (Wendler et al., 2010).[40]

Countries like China and India are "investing substantially in improving their graduate education systems and in the undergraduate programs that feed into those graduate programs" (Wendler et al., 2010, p. 2).[40] Therefore, it is in the best interest of U.S. postbaccalaureate institutions to find out what type of graduate students are enrolling in their degree programs and, most importantly, create a tool that can help with graduate student matriculation and benefit a university by filtering what degree programs are best suited for each type of graduate degree seekers.

In attempts to obtain more knowledge on student matriculation causes, there were many research efforts made to develop student retention theories or matriculation models at the college level. Some of the most notable student retention theories are Tinto's models of student retention (Tinto, 1975),[38] John Bean's model of student attrition (Bean, 1983),[4] and Astin's theory of involvement (Astin, 1984).[2] Based on these theories, researchers have applied different modeling methods to model student matriculation. The most widely used methodologies in the student matriculation research studies are logistic regression (French, Immekus, & Oakes, 2005;[15] Veenstra, Eric, & Gary, 2009),[39] discriminant analysis (Burtner, 2005),[9] and structural equation modeling analysis (Cabrera, Nora, & Castaneda, 1993).[10] Among these previous research efforts, many of them focused mainly on models that predict undergraduate student matriculation. It is still unclear if these models are applicable to choice decisions at the graduate level since these models were developed and designed for use in understanding undergraduate enrollment. This study was designed to build a predictive model for the probability that a specific admitted graduate student will matriculate why Graduate Degrees are Vital to the United States Economy.

Graduate degree holders are crucial to the many challenges that we see in today's innovative world. Recent studies show a shrinking population of graduate students entering the workforce (Allum & Okahana, 2015;[1] Wendler et al., 2010).[40] The U.S. Bureau of Labor Statistics recently projected that employers would add nearly 2.4 million jobs requiring a graduate or even more advanced degree between 2012 and 2022 (Flaherty, 2015).[14] If this projection is correct, then the enrollment of graduate students in the U.S. must increase to meet that demand. Undergraduate education is essential to the creation of a stable economy, providing students with foundational knowledge and the necessary skills that are required to work in a particular field. Meanwhile, graduate education goes beyond basic knowledge; providing students with expertise also further develops critical thinking skills and produces innovators and visionaries (Wendler et al., 2010).[40] Graduate students become modernizers that can lead the way in areas of advancement that are essential to the American economy. Areas such as renewable and alternative energy sources, advanced agricultural practices, pioneering medical procedures, and groundbreaking disease control techniques all require advanced graduate degrees. If the U.S. is to stay advanced in such innovative areas, then appropriate measures need to be taken at the institutional level to help with recruitment and marketing strategies to match the right student with the proper graduate program and degree.

Financial Aid. Inherent in the economic perspective of college enrollment and persistence is the student demand theory, which is related to price. Student demand theory is applicable to enrollment and persistence in higher education as it is a function of the individual's income, the price of an education, the price of the alternatives to a college education, and the aspirations and desires of the individual. Student demand theory assumes that less education will be purchased when the educational costs are high (Leslie & Brinkman, 1987).[23] The impact of financial aid on student matriculation has been widely studied. Although, much of the financial impact occurs at the point of entry, finances may also influence dropout directly through short-term fluctuations in financial need. Terkla and Jackson (1984)[34] found that the cost of education significantly affected students' matriculation decisions, and financial aid awards significantly affected college attendance. However, aid uniquely contributed to the shorter-term measure of academic year persistence.

According to reports by the National Association for College Admission Counseling (NACAC), the 2016 national average acceptance rate at four-year colleges and universities in the United States was 66.1%, up from 63.1% in 2012. Colleges and universities accepted nearly two-thirds of the applicants. The national college six-year graduation rate in the U.S. was 53.8% in 2015 and 55.7 in 2012. During a period of increased acceptance rate in the higher education, decreased percentages of degrees awarded, and increased costs incurred in constant dollars, it seems especially necessary to evaluate the effects of different forms of financial aid on persistence for students presently enrolled in institutions of higher education. National studies analyzing the effects of student financial aid on matriculation have not always concluded that all forms of financial aid had a positive impact on persistence. Therefore, the financial variables in the present study were comprised of assistantship, student loan, and federal student aid.

Different Types of Graduate Students

The potential graduate students who are of importance to the current research are those students who fit into the domestic student category. Domestic students are defined as "citizens or lawful permanent residents of the U.S. or have been granted asylee, refuge or are paroled in the Public Interest status by the U.S. government" (Defining Applicants as International or Domestic, n.d., par.2).[11] These students typically have average GRE scores and have qualified for acceptance into specific programs and universities of their choosing. However, full funding for the domestic student group may not always be available (e.g., full-time and half-time assistantships, full-ride scholarships, and grants). These potential graduate enrollees must assess the cost and benefits of enrolling in a graduate program. The financial risk of increasing the debt that most undergraduate students accrue during their baccalaureate can be a significant setback for most potential graduate students.

Graduate school cost. The Average Student Loan Debt reported that less than 20% of undergraduate students could complete their post-secondary education without accumulating some level of student loan debt (2015).[35] The admission data on domestic students can be broken down and categorized even further by considering the tuition-based cost. While private schools typically cost the same for in-state and out-of-state students, tuition prices at public colleges and universities differ based on this distinction in residency.

The Institute for College Access and Success (2016) [36] reported that post-secondary graduates had incurred an average loan debt of (a) $25,550 from attending a public college, (b) $32,300 from a private university, and (c) $39,950 from a for-profit college. The combination of undergraduate and graduate debt accrued by degree areas breaks down as follows (a) Master of Education is approximately $50,879, (b) Master of Science is approximately $50,400, (c) Master of Arts is $58,539, and (d) law and health sciences range on average between $140,000 - $162,000 (Delisle, Phillips, & van der Linde, 2014).[12] Currently, the most extensive fields of total enrollment were in education and business, with 19% and 15%, respectively, of all graduate students. 73.1% of all students were enrolled in master's degree or graduate certificate programs (Flaherty, 2015).[14]

Undoubtedly the growing cost of attending graduate school has become a problem and a significant setback for most Americans. Most graduate students lessen the financial burden by working full-time while attending graduate school. However, the time and effort that is needed to withstand the demands of graduate school and to keep up with one's personal life leave a minimal amount of time for individuals to do anything else.

In-state residency/Out-of-state. Since each state controls its own education system, each public educational institute is funded by taxpayers' taxes. As a result, in-state residents can attend publicly funded colleges and universities for less cost than out-of-state students/residents.

Predicting Graduate Student Enrollment

Graduate student matriculation can be increased by changing the applicant pool size, employing the right marketing strategies for applicants, and examining admission standards (Khajuria, 2007).[21] As previously discussed, universities have already taken measures to increase their applicant pool size. Therefore this study aims to develop an appropriate model to predict graduate student matriculation based on students' admission data. In doing so, university administrators can employ the correct marketing strategies to attract the right kind of potential graduate students. By utilizing an appropriate prediction model, the university enrollment and recruitment managers can direct their efforts towards prospective graduate students. These are the students that with the correct prediction model can be identified early on and informed of the many possibilities and specific graduate programs that fit their career goals.

Variables that can be used to predict undergraduate student enrollment in a given model may include gender, age, race, social, economic status, residency status, extracurricular activities, high school grade point average (GPA), and SAT/ACT scores. However, some of these variables change at the graduate level. Therefore, the authors built a more appropriate variable selection criteria for graduate student enrollment by reviewing the available literature from comparable resources available (e.g., Council of Graduate School (CGS) reports, the Graduate

DOI: https://doi.org/10.30564/jiep.v4i1&2.2628

Record Examination (GRE) Survey reports, undergraduate matriculation variables, and related case study reports on matriculation prediction).

Graduate student variables

Some of the variables included are non-cognitive variables such as referral data, marital status, and some behavioral data. Data collected on how a potential graduate student was referred to a particular program at the university such as through alumni, friend, or family member is known as referral data. Behavioral data includes the number of days between the application date and admission decision date, if the student confirmed admission (yes/no), and the interactions tracked in Customer Relationship Management (CRM). This study examined and used the non-cognitive variables discussed above along with some cognitive variables such as a graduate student's GRE score, baccalaureate GPA, admission status (regular/ conditional/ provisional), the cost of the graduate program, and whether or not they received their undergraduate degree from the same university. Goenner and Pauls (2006)[17] stated that the selection of variables "differs across models, but generally reflects the academic and personal characteristics of the student in addition to variables that reflect the fit of the institution with a student's preferences" (p. 939). By controlling for such variables, a predictive model can estimate the probability that a potential student who has either applied or been admitted into a specific graduate program will matriculate (Geonner & Pauls, 2006).[17]

Existing Empirical Literature

Most of the research studies on college student matriculation consider the influence of background variables, including ethnicity, age, gender, residency, campus, and location, on student college choice and persistence. Research revealed that the distance between the permanent residence of students and students attending college had a significant direct effect on educational attainment (Bogard, Helbig, Huff, & James, 2011).[5] Pritchard, Klumpp, and Teichler (2015)[26] found that for every percent increase in the distance between the students' home and their college attended, the likelihood of obtaining a degree was decreased by 2.38 percent. A recent study conducted by Garza and Fullerton (2018) [16] found that students who attend colleges at a greater distance from home to school reported a lower level of persistence. Regarding ethnicity, being of African-American descents is included as influential in the persistence process because of a growing concern in higher education that blacks are underrepresented (St. John, Kirshstein, & Noell, 1991).[32] The existing empirical literature focused mainly on models that predict undergraduate student matriculation, although there have been some advancements made in the medical field when it comes to the use of admissions data in predicting medical student matriculation. For instance, in a longitudinal study conducted by Burkhardt, DesJardins, Teener, Gay, and Santen (2016),[8] data was collected from 2006 through 2014 from the University of Michigan Medical School (U-M) and the American Medical College Application Service. The databases were combined to include each applicant's demographic characteristics, along with their academic application scores, institutional financial aid offer, and choice of school to attend. Binomial and multinomial logistic regression models were produced to estimate the predicting factors related to student matriculation at the local institution (i.e., U-M) compared to other highly competitive educational institutes. Both types of logistical models utilized (the binomial and multinomial) were found to be statistically significant ($\rho < .001$) with similar predictive performances (Burkhardt et al., 2016).[8] Results from the binomial model indicated that females, underrepresented minority students, GPA, Medical College Admission Test score, admissions committee desirability score, and most individual financial aid offers were statistically significant ($\rho < .05$) predictors of student matriculation. The multinomial model (excluding females) produced separate likelihoods of students enrolling at different institutional types (Burkhardt et al., 2016).[8] The importance of what should be noted from this study is how imperative it is to have tailored predictive models (i.e., the relevant predictive variables for a given sample) for a university's enrollment management. Predictive models tested and designed to meet the demands of enrollment management and recruitment efforts at a given education institute are crucial to the survival of graduate programs.

Another study conducted by Jeffe and Andriole (2011) [20] focused on the matriculation of MD-Ph.D. graduate students with and without the Medical Scientist Training Program (MSTP) funding. The study examined "the extent to which differences in educational outcomes and career plans exist among MD-Ph.D. program graduates of medical schools with MSTP funding" compared to those without MSTP funding (Jeffe &Andriole, 2011, p. 953).[20] With permission from the Association of American Medical Colleges, the authors examined de-identified records for the national cohorts of all 1993-2000 U.S. medical school matriculants. Jeffe and Andriole (2011) [20] analyzed each MD-Ph.D. graduate student's pre-

 DOI: https://doi.org/10.30564/jiep.v4i1&2.2628

matriculation characteristics, educational outcomes, and career-setting preferences. Three MSTP funded groups were of interest (a) long-standing MSTP funded schools, (b) newly funded MSTP schools, and (c) schools with no MSTP funding. Several multivariate logistic regression models were conducted to test the authors' hypotheses. Jeffe and Andriole (2011)[20] found that the "graduates' pre-matriculation characteristics, educational outcomes, and career plans differed among the three MSTP funding groups" (Jeffe &Andriole, 2011, p. 955).[20]

The authors also concluded that women and nonwhite graduates were more likely to graduate from long-standing MSTP-funded schools. Jeffe and Andriole (2011)[20] also found that while controlling for MSTP school funding, MD-Ph.D. graduates with a total debt of $100,000 or more were more likely to be interested in a non-research-related career (e.g., clinical practice, medical/health career administration, etc.).

Lastly, a study conducted at Howard College of Dentistry by Henley to determine appropriate predictor variables for incoming freshman matriculation utilized several statistical models (viz. factor analysis, discriminate analysis, and logistic regression) to determine the most suitable predictor variables. The study used entrance data collected from the Dental Admissions Test (DAT) (i.e., DAT testing and academic performance data), and all eligible sophomore students' freshman GPA between 2004 and 2008 predict successful graduates. The study's final results indicated a "significant statistical difference in DAT and GPA averages existed between successful and failed students" (Henley, n.d., p. 1).[18] The author's hypothesis on the prediction of failed to successful students was tested via a logistic regression model that resulted in an overall percentage of 87% correct. Therefore Henley's findings suggested that a student's age, DAT score, and freshman GPA do have an overall effect on the student's success in graduating. These results indicate how vital recruitment and targeting the right potential graduate student is for obtaining higher matriculation rates. Enrollment management based approach to recruiting students into specific graduate programs would improve the likelihood of a student's success. It also plays a crucial role in understanding the failed matriculation of highly desirable candidates as well (Burkhardt et al., 2016).[8]

Logistic regression modeling to predict graduate student matriculation

Today's postbaccalaureate applicants are more knowledgeable than those in previous years. They can access information for a specific graduate program, in a given college, at any time. Potential candidates are able to apply to graduate programs by utilizing online formats and filling out online applications, as well as applying for financial aid, assistantships, grants, and scholarships. These students are "shopping for the best package" (Bohannon, 2007, p.1).[6] Hence there is a need for a predictive model that can help assist college administrators and recruiters in reaching the right applicants.

Many of the same factors that influence the decision to apply to a particular graduate program are also the same factors that affect a student's decision to enroll. Predictive models of enrollment are usually based on if a student enrolls or not while conditioning on that student's application to a program or their acceptance (Bruggink & Gambhir, 1996;[7] DesJardins, 2002; [13] Goenner & Pauls, 2006;[17] Leppel, 1993;[22] Thomas, Dawes, & Reznik, 2001).[37] In such cases, where a binary outcome is desired (e.g., matriculation, non-matriculation), the use of a logistic model is appropriate to control for confounding variables. This study employed a logistic regression model to predict graduate student matriculation. Predictive modeling "analyzes past data to make future predictions, or in econometric terms, data is analyzed to estimate a model which is used to make out-of-sample predictions" (Goenner & Pauls, 2006, p. 936).[17] The purpose of this predictive enrollment model is to better understand the contributing characteristics of the incoming graduate student population that help define the variables, which influence a graduate student's decision to matriculate. However, due to the dearth of literature available on graduate student matriculation, selecting the relevant variables for the predictive model was challenging.

2. Methodology

Population

The subjects of this study were 3,718 domestic graduate students who applied and were accepted for admission at an American university over a period of 4 years from 2013 to 2016. The dataset was requested from the Information Management and Technology (IM&T) Department at the university, and it included two subsets. The first one, the development subset, was comprised of 2,573 students who applied and who were admitted during the years from 2013 to 2015. The second one, the validation subset, included 1,145 students who applied and who were admitted at the beginning of 2016.

Research Design

This study was designed to predict the probability that a graduate student who was accepted for admission would actually enroll. No experimental design was necessary

DOI: https://doi.org/10.30564/jiep.v4i1&2.2628

due to the use of historical data. Due to the dichotomous nature of the dependent variable (matriculated/non-matriculated), we applied a logistic regression model which can be used to predict a binary response from multiple independent variables.

The actual enrollment of a student was the dependent variable in this study. When collecting data, the researchers extracted information that indicated whether a particular student enrolled at the university. The dependent variable was coded 1 if the student enrolled; 0 if the student did not enroll, therefore making it binary. Based on previous literature about student matriculation studies and practical feasibilities, a total of 14 predictive variables were included in the logistic regression model. These variables were grouped into four categories: demographic (e.g., age, gender, residency, campus, student's distance to site, and race), academic (e.g., degree level, college, and GPA), financial aid (e.g., assistantship, loan, and FASFA), and behavioral (e.g., time between application data and admission date, experience with the university, and how did you know about the university). The logistic regression model was specified as:

$$\ln(ODDS) = \alpha + \beta_1 \chi_{1i} + \beta_2 \chi_{2i} \ldots . + \beta_n \chi_{ni} + \varepsilon_i \quad (1)$$

$$ODDS = e^{\beta_0 + \beta_1 \chi_{1i} + \beta_2 \chi_{2i} + \ldots + \beta_n \chi_{ni}} \quad (2)$$

where $ODDS$ represents the probability of a student enrolling divided by the probability of a student not enrolling; α is the constant of the equation, also known as the intercept; $\beta_1, \beta_2, \ldots, \beta_n$ are the estimated effects for each corresponding χ (independent variable); and ε_i represents a random error term which is logistically distributed. After calculating the odds ratio for each observation by using formula (1) and (2), the researchers then calculated the probability of enrollment through formula (3)

$$P_i = \frac{ODDS}{1 + ODDS} \quad (3)$$

where P_i is the probability that student i will choose to enroll at the institution to which they applied.

The researchers assessed the predictive accuracy of the final model to predict student matriculation by looking at calibration (or reliability) and discrimination (also called resolution or refinement) in the development subset, as well as in the validation subset. Calibration describes how closely the predicted probabilities agree numerically with the actual outcomes. A standard method that has been widely used to assess model calibration is the Hosmer and Lemeshow (2000)[19] goodness of fit test. Discrimination refers to the ability of a model to correctly distinguish between those with and those without the outcome (Prytherch et al., 2005).[27] Discrimination of the model was examined by calculating the area under the receiver operating characteristic (ROC) curves. The ROC curves show the sensitivity (the proportion of truly positive observations which was classified as matriculated) and specificity (the proportion of truly negative observations which was classified as non-matriculated). The area under the curve, summarized by c-index, represents the likelihood that the proposed model will determine that a student who chooses to matriculate will have a higher probability than a student who chooses not to matriculate. The further the curve above the reference line, the more accurate the model. According to Hosmer and Lemeshow (2000),[19] reasonable discrimination is indicated by c-index values of .7 to .8 and good discrimination by values over .8. Since the predicted values for the dependent variable (matriculation status) are probabilities which range from 0 to 1, the classification of the two-matriculation status (predicted matriculated/predicted non-matriculated) depend on a particular cutoff probability value. The selection of the cutoff probability value was based on the field of study and previous literature. According to Sampath, Flagel, and Figueroa (2009),[28] a student with an estimated 35% to 40% chance of enrolling can be treated as a positive indicator of matriculation. Therefore, the value of .4 was selected as the cutoff probability value in this study.

3. Results

Descriptive Statistics

For the purpose of this study, students who matriculated (2413) and non-matriculated (1306) were identified. The average age of the sample was 30 (SD=8.7), with 70.8% female and 29.2% male. Caucasian students dominated the proportion of the sample with 78.9%, followed by Hispanic (8.9%), other (7.1%), African American (2.5%), and Asian (2.4%). In-state students comprised 53.0% of the student body, and 24.5% of respondents were university's former students. Table 1 and Table 2 provide the details of descriptive statistics for all quantitative variables used in this study.

Logistic Regression

Before performing a series of multiple logistic regression models, the multicollinearity was checked. Multicollinearity occurs in regression models when one predictor variable can be predicted (linearly) from the other predictor variables in the model. According to Pallant (2007),[24] the variance tolerance value cannot be less than .10, nor can the variance inflation factor (VIF) value be larger than 10. The results indicated that none of the variables met this criterion. Therefore, it was judged that multicollinearity was not a factor that could influence the predictive model in this study. Pallant (2007)[24] also recommended that correlations between independent variables should be .30 or better. Correlation analysis among all predictors indicated most pairs to be significant (>.30) and strongly correlated as expected.

A logistic regression model to predict graduate student matriculation was developed with all predictors utilizing the development subset (n=2573). The initial results indicated that the independent variables of race, GPA, residency, and how students knew w about the university were not statistically significant. A subsequent logistic regression model was developed, excluding these four insignificant predictive variables. The results revealed that ten predictive variables were statistically significant to produce the following outcome model:

$$\Pr(Enroll) = \frac{\exp(y)}{1 + \exp(y)}$$

With y equaling -0.797-0.002×(time between application date and admission date)+0.299× (college-1)+ 0.0001× (college-2)-0.018× (college-3)-1.032× (college-4)+0.208× (degree level-1)+0.67× (degree level-2)+0.144× (degree level-3)-0.565× (campus)-0.314× (gender)-0.002× (distance to X))+0.062× (age)+1.333× (FAFSA)+6.463× (assistantship)+0.579× (experience with X).

The model explained 40.1% of the total variance in the log odds for student matriculation by the above ten predictive variables (Cox & Snell R^2=.40). The -2Loglikelihood of 1458.05 was significant. The Omnibus Test of Model coefficient Chi-square was 127.66 with eight degrees of freedom. The beta, standardized error, Wald, degree of freedom, significance, and odds ratio for each significant predictor are displayed in Table 2.

Predictive Accuracy

Calibration. Hosmer and Lemeshow goodness of fit test indicated that the predicted probabilities did not deviate from the probabilities aligned with the prediction of the binary distribution, and the model was adequate for analysis (χ^2 (8, n=2573) =17.38, p=.58).

Discrimination. The researchers used a cutoff probability value of .4 and applied the final prediction model to the validation subset resulting in an overall model accuracy of 77.6%. Specifically, 60.4% of applicants that were predicted to matriculate did matriculate, while 17.2% of the applicants that were predicted not to enroll did not enroll. Table 3, below, presents the predicted and actual matriculation of the validation subset. The c-index had a value of .81, which according to Hosmer and Lemeshow (2000),[19] falls into the group of good discrimination.

Discussion

Through the use of a predictive model, this study has attempted to expand understanding of the college choice process. It has added to and strengthened the literature by giving certain enrollment management professionals a tool to use for predicting their own graduate student enrollment. The results indicated that there were ten predictive variables that were statistically significant at the .05 level. Among the ten predictors, the graduate student's matriculation was mostly influenced by the financial aid variables. For instance, the current study found that the odds of matriculating were 640.9 higher for a student with an assistantship than a graduate student without an assistantship. Moreover, the odds of matriculating are 3.8 times higher when a graduate student receives FAFSA. This is consistent with countless studies in student retention (e.g., DesJardins, Ahlburg, & McCall, 2002;[13] Singell, 2004;[29] St. John, 2000),[33] which have illustrated the positive effect financial aid factors have had on student matriculation. For example, conducted a study to investigate the impact of financial aid on student matriculation for public/private higher education. The results demonstrated that financial aid had a significant positive impact on matriculation. Providing adequate funds for students who are unable to defray the full costs of higher education has always perplexed postsecondary institution administrators. These days, postsecondary institutions are no longer in a seller's market. As a result, students' buying habits have also changed. To an increasing degree, graduate students base their initial entry decisions and their staying or leaving decisions on their perception of the cost of attendance. There were eight other significant predictors, including experience with university (former student were 1.8 times more likely to enroll than students who had not attended the university), campus (off-campus students were 1.8 times more likely to enroll than on-campus students), degree level (specialist students were two times more likely to enroll than doctoral students), college (compared with the business major students, students majoring in education were 2.8 times more likely to enroll), gender (males were 1.4 times more likely to enroll than females), age (with one unit increase in age, the odds of matriculating increase by one), the number of days between application and admission (with one unit increase in the number of days between application and admission, the odds of matriculating increase by one), and distance to the university (with one unit increase in distance to the university, the odds of matriculating increase by one).

DOI: https://doi.org/10.30564/jiep.v4i1&2.2628

Practical Implication

The primary purpose of this study was to develop a predictive model that would more accurately predict a graduate student's likelihood to enroll at a university. The predictive model of matriculation generated from this study could be used by the universities to enhance their recruitment efforts. For example, using the current software program, the Department of Student Admission and Recruitment could calculate the predictive score (or percentage) for each prospective student based on the predictive model. The office personnel could then focus their time and monetary resources on those applicants whose c-index was .4 or higher. Furthermore, after running prospective students' information through the predictive model, enrollment managers could then eliminate the low qualifying students from the recruitment plan. This would allow the staff to spend more time on the qualifying graduate students and save the institution financially by not mailing the low qualifying students as much direct mail.

Limitations and Future Research

While a model that targets applicants or admitted students is helpful, it is still limited to a small population of graduate students who have already shown interest in enrolling at one university. A more efficient way of drawing from a larger pool of potential applicants would be to implement a predictive model much early on in the recruitment stage such as in the inquiry stage (i.e., when the student is still searching for information on graduate degree programs). A model as Goenner and Pauls (2006) [17] suggested, "can early on in the recruitment cycle provide admissions and recruitment officers with a tool to craft recruitment and marketing efforts to solicit more applications and increase enrollment" (p. 940). An inquiry may include "a student filling out an information card, attending a college fair, making a campus visit, sending an email, or phoning to request information" (Goenner & Pauls, 2006, p. 937).[17] Focusing on the student's data collected from their inquiry into the university helps to estimate a student's interest early on in the recruitment process, which can target specific marketing and recruitment efforts based on a potential applicant's interest (Goenner & Pauls, 2006).[17] The exchange of information at the inquiry level may be the most important for institutions to increase applications and enrollment (Paulsen, 1990).[25]

References

[1] Allum, J., & Okahana, H. (2015). *Graduate enrollment and degrees: 2004 to 2014* [PDF]. Washington, DC: Council of Graduate Schools.

[2] Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel*, *25*(4), 297-308.

[3] Baum, S., Kurose, C., Mcpherson, M. (2013). An overview of American higher education. *The Future of Children*, 23(1), 17-39.

[4] Bean, J. P. (1983). The application of a model of turnover in work organizations to the student attrition process. *The Review of Higher Education*, *6*(2), 129-148.

[5] Bogard M., Helbig, T., Huff, G., & James, C. (2011). A comparison of empirical models for predicting student retention. Bowling Green: Western Kentucky University.

[6] Bohannon, T. (2007, April 16). *Predictive Modeling in Higher Education*. Speech presented at SAS Global Forum in Walt Disney World Swan and Dolphin Hotel, Orlando, Florida.

[7] Bruggink, T., & Gambhir, V. (1996). Statistical models for college admission and enrollment: A case study for a selective liberal arts college. *Research in Higher Education, 37*(2), 221-240.

[8] Burkhardt, J., DesJardins, S., Teener C., Gay, S., & Santen, S. (2016). Enrollment management in medical school admissions: A novel evidence-based approach at one institution. *Academic Medicine* 91(11) 1561-1567.

[9] Burtner, J. (2005). The use of discriminant analysis to investigate the influence of non-cognitive factors on engineering school persistence. *Journal of Engineering Education*, *94*(3), 335-338.

[10] Cabrera, A. F., Nora, A., & Castaneda, M. B. (1993). College persistence: Structural equations modeling test of an integrated model of student retention. *The Journal of Higher Education*, *64*(2), 123-139.

[11] Defining Applicants as International or Domestic. (n.d.). Retrieved October 15, 2016, from http://www.grad.illinois.edu/admissions/instructions/02b

[12] Delisle, J., Phillips, O., & van der Linde, R. (March, 2014). *The graduate student debt review: The state of graduate student borrowing* [PDF]. New America Education Policy Program.

[13] DesJardins, S. (2002). An analytic strategy to assist institutional recruitment and marketing efforts. *Research in Higher Education, 43*(5): 531-553.

[14] Flaherty, C. (2015, September 17). Going Back to School. *Inside Higher Ed*. Retrieved September 21, 2016, from https://www.insidehighered.com/news/2015/09/17/study-finds-first-time-enrollment-graduate-school-35-percent

[15] French, B. F., Immekus, J. C., & Oakes, W. C. (2005). An examination of indicators of engineering students' success and persistence. *Journal of Engineering Education*, *94*(4), 419-425.

[16] Garza, A.N., & Fullerton, A.S. (2018). Staying lose or going away: how distance to college impacts the educational attainment and academic performance of first-generation college students. *Sociological Perspectives, 61*(1), 164-185.

[17] Goenner, C., & Pauls, K. (2006). A predictive model of inquiry to enrollment. *Research in Higher Education, 47*(8), 935-956.

[18] Henley, G. (n.d.) *Predicting successful matriculation through basic sciences* [PDF]. Retrieved from http://www.gs.howard.edu/middle_states/mse/supportingdocs/chapter%2010%20-%20institutional%20assessment/10.7%20Case%20Study%20-%20Predicting%20Successful%20Matriculation%20through%20Basic%20Sciences.pdf

[19] Hosmer, D., & Lemeshow, S. (2000). *Applied Logistic Regression* (2^{nd} edition). New York: Wiley.

[20] Jeffe, D. & Andriole, D. (2011). A national cohort study of MD-PhD graduates of medical schools with and without funding from the national institute of general medical sciences' medical scientist training program. *Careers in Academic Medicine 86*(8), 953-961.

[21] Khajuria, S (2007). *A model to predict student matriculation from admissions data* (Master's thesis). Retrieved from Ohio University database http://rave.ohiolink.edu/etdc/view?acc_num=ohiou1167852960

[22] Leppel, K. (1993). Logit estimation of a gravity model of the college enrollment decision. *Research in Higher Education, 34* (3), 387-398.

[23] Leslie, L. & Brinkman, P. (1987). Student price response in higher education: The student demand studies. *The Journal of Higher Education*, 58, 181-204.

[24] Pallant, J. (2007). *SPSS survival manual: A step by Step guide to data analysis using SPSS for windows* (Third). New York: MCGraw-Hill Higher Education.

[25] Paulsen, M. B. (1990). *College choice: Understanding student enrollment behavior*. ASHE-ERIC Higher Education Report No.6. Washington, DC: The George Washington University, School of Education and Human Development.

[26] Pritchard, R., Klumpp, M., & Teichler, U. (Ed.). (2015). *Diversity an excellent in higher education*. Springer.

[27] Prytherch, D., Briggs, J., Weaver, P. C., Schmidt, P., & Smith, G. B. (2005). Measuring clinical performance using routinely collected clinical data. *Medical Informatics and the Internet in Medicine, 30*(2), 151-156.

[28] Sampath, V., Flagel, A., & Figueroa, C. (2009, October). *A logistic regression model to predict freshmen enrollments*. Paper presented at the annual meeting of the SouthEast SAS Users Group. Retrieved from http://analytics.ncsu.edu/sesug/2009/SD016.Sampath.pdf [Google Scholar]

[29] Singell, D. (2004). Come and stay a while: does financial aid effect retention conditioned on enrollment at a large public university? *Economics of Education Review, 23*(5), 459-471.

[30] Shanghai Ranking Consultancy (2009). Academic rankings of world universities. Shanghai, China: Shanghai Jiao Tong University. Retrieved from http://www.arwu.org/

[31] Snyder, T.D., de Brey, C., & Dillow, S.A. (2018). *Digest of Education Statistics 2016* (NCES 2017-094). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

[32] St. John, E., Kirshstein, R., and Noell, J. (1991). The impact of student financial aid on persistence: A sequential analysis. *Review of Higher Education, 14*(3): 383–406.

[33] St. John, E. (2000). The impact of student aid on recruitment and retention: What the research indicates. *New Directions for Student Services, 89*, 61-75.

[34] Terkla, D., & Jackson, G. (1984). *Conceptual models of student choice*. Boston: Harvard University.

[35] The Average Student Loan Debt (2015). Retrieved September 30, 2016, from http://www.collegescholarships.org/loans/average-debt.htm

[36] The Institute for College Access and Success. (2016). *Student debt and the class of 2015*. Retrieved from: https://ticas.org/wp-content/uploads/legacy-files/pub_files/classof2015.pdf

[37] Thomas, E., Dawes, W., & Reznik, G. (2001). Using predictive modeling to target student recruitment: Theory and practice. AIR Professional File 78(Winter): 1–8.

[38] Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, *45*(1), 89-125.

[39] Veenstra, C. P., Dey, E. L., & Herrin, G. D. (2009). A Model for Freshman Engineering Retention. *Advances in Engineering Education*, *1*(3), 1-33.

[40] Wendler, C., Bridgeman, B., Cline, F., Millett, C., Rock, J., Bell, N., & McAllister, P. (2010). The Path Forward: The Future of Graduate Education in the United States. Princeton, NJ: Educational Testing Service

  DOI: https://doi.org/10.30564/jiep.v4i1&2.2628

Table 1 Descriptive Statistic for Development Subset

Categorical Variables	n(Matriculated)	%(Matriculated)
Gender		
Male(n=773)	520	67.3%
Female(n=1800)	1170	65.0%
College		
EBS(n=1087)	798	73.4%
NHS(n=1023)	582	56.9%
PVA(n=305)	203	66.0%
HSS(n=119)	79	66.4%
MCB(n=39)	28	71.8%
Campus		
Main(n=1495)	829	55.5%
Off(n=1071)	858	80.1%
Degree Level		
Doctoral(n=485)	311	64.1%
Master(n=1833)	1218	66.4%
Specialist(n=147)	92	62.6%
Certificate(n=108)	69	63.9%
Assistantship		
Yes(n=321)	320	99.7%
No(n=2252)	1370	60.8%
Race		
White(n=2044)	1353	66.2%
Hispanic(n=214)	152	71.0%
Black/Afr. Am.(n=62)	43	69.4%
Asian(n=57)	36	63.2%
Other(n=196)	106	54.1%
Experience with University		
Yes(n=657)	542	82.5%
No(n=1916)	1148	59.9%
Residency		
In-State(n=1390)	1102	79.3%
Out-State(n=1182)	587	49.7%
FAFSA		
Yes(n=1583)	1210	76.4%
No(n=990)	480	48.5%

	Min.	Max.	M	SD
Continuous Independent variables				
Distance to University (Miles)	.1	3367	34	504.1
# of days from application - admission	2	808	97	87.6
Age	20	67	30	8.9
GPA	1.9	4.0	3.6	.3

Note: M=Mean, SD= Standardized deviation

Table 2 Descriptive Statistic for Validation Subset

Categorical Variables	n(Matriculated)	%(Matriculated)
Gender		
Male(n=315)	520	67.3%
Female(n=830)	1170	65.0%
College		
EBS(n=447)	798	73.4%
NHS(n=529)	582	56.9%
PVA(n=109)	203	66.0%
HSS(n=42)	79	66.4%
MCB(n=18)	28	71.8%
Campus		
Main(n=627)	829	55.5%
Off(n=518)	858	80.1%
Degree Level		
Doctoral(n=193)	311	64.1%
Master(n=872)	1218	66.4%
Specialist(n=39)	92	62.6%
Certificate(n=41)	69	63.9%
Assistantship		
Yes(n=149)	320	99.7%
No(n=996)	1370	60.8%
Race		
White(n=891)	1353	66.2%
Hispanic(n=119)	152	71.0%
Black/Afr. Am.(n=31)	43	69.4%
Asian(n=34)	36	63.2%
Other(n=70)	106	54.1%
Experience with University		
Yes(n=253)	542	82.5%
No(n=892)	1148	59.9%
Residency		
In-State(n=581)	1102	79.3%
Out-State(n=564)	587	49.7%
FAFSA		
Yes(n=706)	1210	76.4%
No(n=493)	480	48.5%

	Min.	Max.	M	SD
Continuous Variables				
Distance to University (Miles)	.1	3369	424	535.9
# of days from application - admission	4	1201	92	92.4
Age	19	45	30	8.6
GPA	2.1	4.1	3.6	.4

Note: M=Mean, SD= Standardized deviation

 DOI: https://doi.org/10.30564/jiep.v4i1&2.2628

Table 3 Logistic Regression Results

Predictor	B	SE	Wald	df	Sig.	OR
Time between application date and admission date	-.002	.001	3.903	1	.048	.998
College			15.337	1	.004	
College (1)	.299	.432	.479	4	.489	1.348
College (2)	.001	.429	.001	1	.999	1.000
College (3)	-.018	.473	.002	1	.969	.982
College (4)	-1.032	.520	3.935	1	.047	.356
Degree level			9.255	1	.026	
Degree level (1)	.208	.379	.300	3	.584	1.231
Degree level (2)	.670	.345	3.780	1	.049	1.954
Degree level (3)	.144	.431	.111	1	.739	1.154
Campus	-.565	.183	9.577	1	.002	.568
Gender	-.314	.146	4.637	1	.031	.730
Distance to University	-.002	.001	68.193	1	<.001	.998
Age	.062	.011	31.349	1	<.001	1.064
FAFSA	1.333	.142	88.687	1	<.001	3.791
Assistantship	6.463	1.032	39.207	1	<.001	640.851
Experience with University	.579	.177	10.723	1	.001	1.785
Constant	-.797	1.038	.589	1	.443	.451

Note: B=beta weight, SE=Standard error, df = Degree of freedom; Sig.= Significance; OR = Odds ratio

Table 4 Predicted and Actual Matriculation of Validation Subset

Matriculated		Predicted	
		Non-Matriculated	
Actual	Matriculated	692 60.4%	31 2.7%
	Non-Matriculated	225 19.7%	197 17.2%

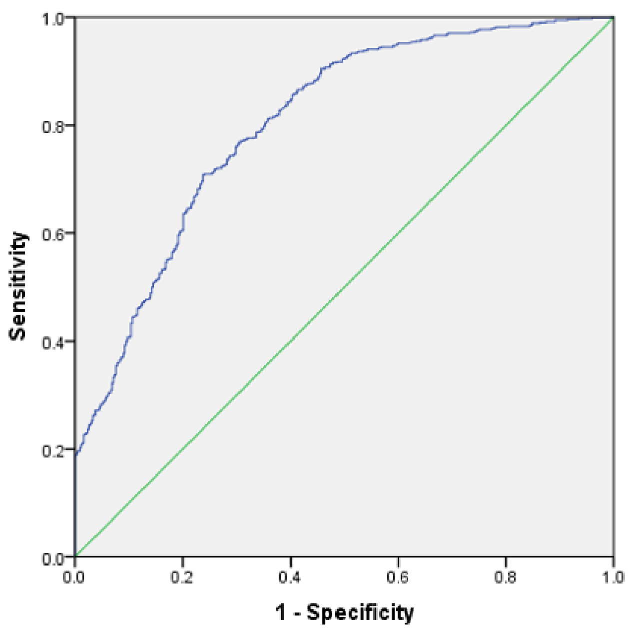

Figure 1 Receiver Operating Characteristic Curve

DOI: https://doi.org/10.30564/jiep.v4i1&2.2628